The Crystal Structure of Mismatch-specific Uracil-DNA Glycosylase (MUG) from *Deinococcus radiodurans* Reveals a Novel Catalytic Residue and Broad Substrate Specificity*

Elin Moe†, Ingar Leiros§, Arne O. Smalås¶, and Sean McSweeney∥

From the †Norwegian Structural Biology Centre, University of Tromsø, N-9037 Tromsø, Norway and the §European Synchrotron Radiation Facility, BP 220, F-38043 Grenoble Cedex 9, France

*This work was financed by the Functional Genomics program (FUGE) of the Norwegian Research Council and the European Synchrotron Radiation Facility in Grenoble, France. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Deinococcus radiodurans is extremely resistant to the effects of ionizing radiation. The source of the radiation resistance is not known, but an expansion of specific protein families related to stress response and damage control has been observed. DNA repair enzymes are among the expanded protein families in *D. radiodurans*, and genes encoding five different uracil-DNA glycosylases are identified in the genome. Here we report the three-dimensional structure of the mismatch-specific uracil-DNA glycosylase (MUG) from *D. radiodurans* (drMUG) to a resolution of 1.75 Å. Structural analyses suggest that drMUG possesses a novel catalytic residue, Asp-93. Activity measurements show that drMUG has a modified and broadened substrate specificity compared with *Escherichia coli* MUG. The importance of Asp-93 for activity was confirmed by structural analysis and abolished activity for the mutant drMUGD93A. Two other microorganisms, *Bradyrhizobium japonicum* and *Rhodopseudomonas palustris*, possess genes that encode MUGs with the highest sequence identity to drMUG among all of the bacterial MUGs examined. A phylogenetic analysis indicates that these three MUGs form a new MUG/thymidine-DNA glycosylase subfamily, here called the MUG2 family. We suggest that the novel catalytic residue (Asp-93) has evolved to provide drMUG with broad substrate specificity to increase the DNA repair repertoire of *D. radiodurans*.

Uracil-DNA glycosylases are responsible for the removal of uracil from DNA as part of the base excision repair pathway. Uracil in DNA may result from misincorporation during replication or by deamination of cytosine resulting in A:U and G:U mismatches (1). Six members of the uracil-DNA glycosylase (UDG) family have been identified as follows: (i) uracil-DNA N-glycosylase (UNG) (1); (ii) mismatch-specific DNA-glycosylases (2, 3); (iii) single-strand selective monofunctional UDG (4); (iv) thermostable UDG (5); (v) UDG-B (6, 7); and (vi) MIG protein/endonuclease III/Methanococcus jannaschii UDG family (6, 7).

Thymine DNA-glycosylase (TDG) removes uracil and thymine from G:U and G:T mismatches in double-stranded DNA (dsDNA) (8). Mismatch-specific uracil-DNA glycosylase (MUG) is the prokaryotic homologue of TDG, and *Escherichia coli* MUG (ecMUG) has been shown to be highly specific to G:U mismatches but also to repair G:T mismatches at very high enzyme concentrations. The crystal structure of ecMUG revealed that it consists of a central 5-stranded β-sheet flanked on both sides by α-helices and that ecMUG and human uracil-DNA N-glycosylase (hUNG) are structurally related (9). Some of the same structural elements in hUNG and ecMUG are involved in catalysis; however, the catalytically important residues are different.

The highly conserved [GQDPYH]148 sequence in the so-called water-activating loop in hUNG is substituted by [GINPG]20 in ecMUG (a conserved sequence motif among most MUG/TDGs), and the catalytic residue Asp-145 is substituted by Asn-18. In addition, the [HPSPLS]273 sequence in the highly conserved DNA intercalating loop (Leu-272 loop) in hUNG is substituted by [NPSGLS]146 in ecMUG, and the catalytically important His-268 in hUNG is replaced by Asn-140 (9). Leu-272 in hUNG is observed to fill the gap created in the DNA duplex when uracil is flipped out of the DNA strand and into the substrate-binding pocket (10). ecMUG contains a Leu residue (Leu-144) in a topologically similar position, and this residue is believed to have the same function (9). In addition it has been shown that the neighboring residues Ser and Arg, the latter in particular, intercalate with the widened guanine on the complementary DNA strand, and the contacts formed were thus thought to be specific in recognizing the guanine opposite of the mismatched uracil instead of uracil itself. Furthermore, ecMUG was found to be highly specific for dsDNA, as it does not excise uracil from single-stranded DNA (9).

hUNG is suggested to perform the catalysis of the N-glycoside bond in a stepwise, dissociative manner. The cleavage of the glycosyl bond yields an intermediate that comprises an oxocarbenium cation and an uracil anion. A subsequent attack by a water molecule and transfer of a proton to Asp-145 results in the products. The side chains of both Asp-145 and His-268 contribute to lowering the activation energy of the reaction by stabilizing the transition state. Asp-145 has been suggested to participate in stabilizing the oxocarbenium cation, whereas His-268 is thought to participate in the stabilization of the uracil anion (11). In ecMUG, both Asp-145 and His-268 are substituted by asparagines (Asn-18 and Asn-140). Asn-18 in MUG, which is conserved in most known MUG/TDG sequences, binds and presents a water molecule in a similar position and orientation to the catalytic aspartate of hUNG and is aligned for nucleophilic attack on the deoxyribose in order for the catalysis to take place. Although there is a possibility that Asn-18 could participate in stabilizing the transition state in the MUG/TDG class of enzymes also, the residue corresponding to Asn-140 in ecMUG is poorly conserved throughout the MUG/TGD class of enzymes. There-
Mismatch-specific Uracil-DNA Glycosylase from D. radiodurans

fore, it has not been verified that the MUG/TDG enzymes follow the stepwise dissociative reaction pathway that is now commonly accepted for UNG. Even though some of the catalytic power of hUNG is retained in ecMUG, it is not able to perform the same efficient catalysis because of the lack of the enzymatic properties of Asp and His. This probably explains the observed low rate of uracil excision of ecMUG compared with hUNG (9).

Deinococcus radiodurans is a non-pathogenic soil bacterium well known for its extreme ability to withstand ultraviolet and ionizing radiation as well as desiccation. D. radiodurans tolerates ionizing radiation at doses lethal to other organisms and is capable of surviving 5,000–30,000 grays of ionizing radiation (12), whereas most other organisms cannot survive doses of above 50 grays. Such a massive radiation dose is estimated to induce several hundred double-strand breaks, thousands of single-strand gaps, and ~1,000 sites of DNA base damage per chromosome (13). In the case of acute radiation of ~1,000 grays, D. radiodurans repairs its DNA efficiently within a few hours, whereas other organisms typically cannot tolerate more than two to three DNA radiation-induced double-strand breaks per chromosome (14). The fact that D. radiodurans is so resistant to radiation has made it an attractive target for studies on bio-remediation of radioactivity and/or chemically contaminated sites, and mainly for that reason the D. radiodurans genome was sequenced (15). Despite thorough biological studies of this organism, the mechanisms of the radiation resistance remains unclear; however an expansion of specific protein families related to stress response, damage control, and DNA repair is observed in D. radiodurans (16).

DNA glycosylases are among the expanded protein families in D. radiodurans, and for the UDG family five UDG-encoding genes have been identified (16). Here we report the crystal structure of D. radiodurans MUG (drMUG/DR0715) to 1.75-Å resolution and that of the mutant drMUGD93A to 1.7-Å resolution. The overall structure of drMUG is similar to that of E. coli MUG; however, drMUG possesses a novel catalytic residue (Asp-93) and has broader substrate specificity compared with E. coli MUG.

EXPERIMENTAL PROCEDURES

Cloning—The gene encoding the G:U mismatch uracil-DNA glycosylase from D. radiodurans, dr0715, was inserted into the pDEST14 expression vector using the Gateway® technology according to the manufacturer’s manual (Invitrogen). The primers used for amplification of the gene (Sigma Genosys) were the forward drMUG primer (5’-CATCACCATCACATCACTGCGCCGCGCATGACGTG-3’), the reverse drMUG primer (5’-GGGGACCACTTTGTACAAGAAAGCGGTTGTCTACATGACGGGGGTTTCCGGG-3’), and the forward D. radiodurans His tag primer 5’-GGGCAACAGTGTACTACAAAAGGTCAGGCCTTCAGATAGAACCATGCATCACCATCACCATCACCTC-3’. The dr071S gene was first amplified using the forward drMUG and reverse drMUG primers and thereafter using the forward D. radiodurans His tag and reverse drMUG primers in order to engineer 18 additional nucleotides encoding an N-terminal hexahistidine tag. The resulting PCR product was used in the BP reaction with the pDONR201 vector and in the LR reaction with the pDEST14 vector. The sequence of the clone was confirmed by DNA sequencing (ABI 377 DNA sequencer, Amersham Biosciences). The drMUGD93A mutant was constructed using the QuikChange® site-directed mutagenesis kit following the manual from the manufacturer (Stratagene). The primers used were the OPD93A primer (5’-GCCGCCACAGGGGGTTGGGGGCCGCACTC-3’) and the NPD93A primer (5’-GAGTGGCGCCGCGC-}

FIGURE 1. SDS-PAGE analysis of purified drMUG and drMUGD93A (1–3 μg of each protein). Mark 12345 (Invitrogen) unstained protein standard was used as the protein standard (protein size given in kDa). Lane 1, drMUG; lane 2, drMUGD93A.

CACCCTCGTGTGGCGCGC-3’). The mutation was confirmed by DNA sequencing.

Expression and Purification—The BL21(DE3)pLysS expression strain (Stratagene) was transformed with the pDEST14 vector containing dr0715 and dr0715D93A, respectively, and expression was induced for 3 h after the addition of 0.5 mM isopropyl-β-d-thiogalactopyranoside at 37 °C in 1 liter of Luria-Bertani medium containing 100 μg/ml ampicillin. The soluble fraction of the lysed cells was loaded onto a 5-ml HiTrap HP chelating Sepharose column charged with 100 mM NiSO₄ equilibrated with 50 mM Tris/HCl, pH 7.5, and 150 mM NaCl. The proteins were eluted from the column using a gradient from 0–100% 50 mM Tris/HCl, pH 7.5, 150 mM NaCl, and 500 mM imidazole and contained ~200 mM imidazole. The proteins were thereafter concentrated to 10 mg/ml (drMUG) and 43 mg/ml (drMUGD93A) and stored at 4 °C.

Protein Determination and Analysis—Protein concentrations were determined with Bio-Rad protein assay dye reagent concentrate based on the Bradford dye binding procedure (17) (Bio-Rad) using bovine serum albumin as standard. Protein purity was determined using SDS-PAGE (NuPAGE) with 4–12% Bis-Tris gels run in MET buffer. The gel was stained in Simply Blue SafeStain according to the protocol supplied by the manufacturer (Invitrogen).

Specificity Measurements—Substrate specificity was measured using 30-bp dsDNA oligonucleotide substrates containing G-C, G-U, G-T, and A-U base pairs and one single-stranded DNA containing uracil labeled with fluorescein in the 5’-end (MWG Biotech AG). The dsDNA oligonucleotide had the sequence 5’-CCACCGGCTATCGTGTTGCGCGC-3’ (18) where the boldfaced X was either C, U, or T; in the reverse and the complementary strands, G or A was a partner to the X. Reactions between protein and DNA were performed in 20 mM Tris/ HCl, pH 8.0, 0.1 mg/ml bovine serum albumin, 1 mM EDTA, 1 mM EGTA, 1 mM dithiothreitol, 1 pmol of DNA, 1 pmol of drMUG and drMUGD93A, and 1 unit of ecMUG (Nordic Biosite AB) in a 20-μl reaction mixture for 1 h at 37 °C. The reaction was stopped, and the resulting abasic site was cleaved off by adding NaOH to a final concentration of 100 mM and incubating at 99 °C for 10 min; thereafter, it was analyzed on a denaturing 15% Tris borate-EDTA/urea PAGE (Invitrogen). The products were visualized using the UV light illumination of the Gel Doc 2000 system (Bio-Rad).
Specific activity was measured using nick-translated calf thymus DNA (Sigma-Aldrich) with deoxy[5-3H]uridine 5'-triphosphate (Amersham Biosciences). Preparation of substrate and measurement of activity were performed as described previously (19).

**Crystallization and Data Collection**—Crystallization was performed using the hanging drop method. Crystals of recombinant drMUG and drMUGD93A were grown by mixing 1-μl drops of 10 mg/ml protein with a solution containing 0.2 M sodium acetate trihydrate, 0.1 M sodium cacodylate at pH 6.5, and 30% (w/v) polyethylene glycol 8000 (Crystal screen I, number 28; Hampton Research). The drops were equilibrated at 18 °C, and hexagonal crystals suitable for data collection purposes appeared after 1 day. The crystals had overall dimensions of about 80 × 80 × 80 μm³ and were flash-frozen directly from the drop using a nitrogen cold stream (Oxford Instruments) operating at a temperature of 100 K. All diffraction data were collected at the European Synchrotron Radiation Facility (Grenoble, France), macromolecular crystallography beamline ID14-EH4 (native) and ID14-EH1 (mutant). Complete data sets (see Table 2) were collected to a resolution of 1.75 and 1.7 Å for the native and mutant drMUGs, respectively. These data allowed the solution of both crystal structures using molecular replacement techniques.

**Structure Determination and Refinement**—The data sets collected were indexed and integrated using MOSFLM (20). The unit cell parameters of drMUG were \(a = b = 101.7\,\text{Å}, c = 37.6\,\text{Å}, \alpha = \beta = 90^\circ,\) and \(\gamma = 120^\circ\), and for drMUGD93A the parameters were \(a = b = 102.5\,\text{Å}, c = 37.6\,\text{Å}, \alpha = \beta = 90^\circ,\) and \(\gamma = 120^\circ\); both belong to the hexagonal space group.
group P6. The data were scaled and merged, and the intensities were converted into structure factors using the CCP4 programs SCALA and TRUNCATE (21). A summary of the data collection statistics is presented in Table 2. It was assumed that one molecule in the asymmetric unit leads to a solvent content of roughly 52%, with a Matthew’s coefficient of 2.6 Å³ Da⁻¹. Molecular replacement was performed using MOLREP (21). The crystal structure of ecMUG (9) was used as an initial model for drMUG, and the automated program functions in MOLREP were applied to create the search model that presumably had the best fit to the sequence of drMUG. Reflections to a high resolution limit of 4 Å were used, and one well resolved solution could be found.

Rigid body fitting of the model using a high-resolution cutoff at 3.5Å resulted in an \( R_{work} \) of 48.9% (\( R_{free} \) of 49.6%). Automated model building with ARP/wARP (22), including all reflections to 1.75 Å, built a total of 177 amino acid residues in two chains into electron density. After manual intervention using O (23) the model was refined in REFMAC5 (24), resulting in \( R_{work} \)-factors of 23.5 and 27.1% for the working and test sets of reflections, respectively, including all reflections in the refinement. Subsequent cycles of refinement interspersed with manual rebuilding gave final \( R_{work} \) and \( R_{free} \) values of 17.9 and 22.0%, respectively, with acceptable geometry. The crystal structure of the drMUGD93A mutant was determined by molecular replacement with MOLREP (21) using the refined structure of drMUG as a search model with water molecules, acetate, and Asp-93 removed. The model was refined as for drMUG with final \( R_{work} \) and \( R_{free} \) values of 20.8 and 24.4%, respectively, after refinement. For an overview of the refinement statistics, see Table 2.

RESULTS AND DISCUSSION

Protein Expression and Purification—The expression of the protein and its purification to apparent purity were confirmed by SDS-PAGE as shown in Fig. 1. The molecular mass of the purified protein seems to be larger than the theoretical mass 21.8 kDa, as judged from electrophoresis with the SDS-polyacrylamide gel.

The Overall Structure of drMUG—The refined model of drMUG comprises a single polypeptide chain where amino acid residues 7–189 are defined in electron density. No electron density was observed for the N-terminal residues 1–6 and the C-terminal residues 189–199. Two short loop regions in the structure were poorly defined in the initial electron density maps (residues 144–147 and 165–167) but could be built and were included in the subsequent refinement cycles. One acetate and 151 water molecules were also incorporated and used in refinement.

The central part of the structure is a five-stranded \( \beta \)-sheet having a strand order of 2-1-3-5-4, with strand 4 antiparallel to the others. This \( \beta \)-sheet is flanked on both sides by \( \alpha \)-helices (Fig. 2A). Overall, the crystal structure of drMUG is similar to the previously determined structure of ecMUG (Protein Data Bank code 1MUG), with a
root mean square deviation of 1.15 Å for 146 aligned Cα-atoms (Fig. 2B). A pronounced positively charged cavity on the C-terminal side of the β-sheet has been shown to be contacting the damaged DNA in ecMUG (9). DrMUG can also be structurally aligned with the crystal structure of unbound hUNG (Protein Data Bank code 2SSP) (25) with an root mean square deviation of 1.74 Å for 64 Cα-atoms. The most similar residues are as expected in the central β-sheet and in the DNA contacting region.

A planar, Y-shaped electron density was situated in the substrate-binding site. This moiety was modeled as an acetate molecule, consistent with the high concentrations of acetate used in the crystallization of drMUG. The acetate is tightly coordinated by Asp-84 in the bottom of the substrate-binding pocket (Fig. 3). Compared with the crystal structure of ecMUG complexed with the non-hydrolyzable deoxyuridine analogue β-FU (26), the acetate in drMUG is situated much deeper in the substrate-binding pocket (not shown). This tight interaction is made possible by the substitution of Lys-68 in ecMUG by Asp-84 in drMUG. It has further been suggested that ecMUG possesses a so-called “back door” escape route for excised bases, as the bottom of the substrate-binding pocket is open and connects to the surface close to the N terminus of the protein (26). In drMUG this hypothesized escape route is closed because it has 16 additional N-terminal residues as compared with ecMUG (Fig. 4). In the crystal structure, the longer N-terminal part of drMUG forms a lid that eliminates the possibility for a cleaved base to escape via this route (Fig. 2B).

**Nucleotide Stabilization and Discrimination**—To perform efficient glycoside bond cleavage, drMUG must stabilize the mismatched uracil in the specificity pocket. Additional mechanisms should be in place to allow discrimination against thymine and cytosine. The nucleotide stabilization is most likely provided by Tyr-46 in drMUG (Fig. 4). An analysis of the sequences of MUGs from various organisms shows that Tyr-46 is conserved in all MUGs, indicating its importance in nucleotide discrimination.
Mismatch-specific Uracil-DNA Glycosylase from *D. radiodurans*

ysis of the crystal structure shows that this residue will make stacking interaction with the flipped-out nucleotide in the specificity pocket (Fig. 5). A sequence alignment of DrMUG, ecMUG, and three UNG sequences reveals that the equivalent residues in hUNG and ecMUG is Phe-158 and Phe-30, respectively (Fig. 4). A superimposition of the specificity pockets of DrMUG and ecMUG shows how Tyr-46 in DrMUG and Phe-30 in ecMUG can stack against the nucleotide (Fig. 5).

A ClustalW alignment (27) of the 15 sequences with the highest identity to DrMUG was made in ExPASy (28). From the alignment, it is clear that DrMUG is rather similar to the eukaryotic TDGs (Fig. 6). It is also apparent that the amino acid in position 46 is strictly conserved as an aromatic residue throughout the MUG/TDG family. This aromatic interaction thus appears to be conserved among UNG and MUG/TDG enzymes as a mechanism of substrate stabilization in the specificity pocket.

Binding of thymine in the activity pocket is probably prevented by Ser-36 and Ser-39 in DrMUG (Fig. 4). The crystal structure suggests that the side chains of both these two Ser residues would prevent thymine from binding in the active site of DrMUG by a steric hindrance of the methyl group (Fig. 5). In ecMUG Ser-36 is substituted by Gly-20, whereas Ser-39 is conserved as Ser-23 (Fig. 4). In a structure of ecMUG-β-FU DNA, one can see that the C5 position of uracil is directed toward the side chain of Ser-23 in such a way that it will sterically hinder the C5-methylated base in thymine (26) (Fig. 5). As the side chain of Ser-23 can rotate, it will probably not totally exclude the binding of thymine in the active site, which probably explains the ability of ecMUG to excise G:T mismatches at an elevated enzyme concentration (9). However, the G:T processing activity of ecMUG has been suggested not to be of any biological significance because of its low kinetic constant compared with other mismatches (29). In a similar manner, the binding of thymine in hUNG is hindered by the side chain of Tyr-147 in the 143GQDPY147 sequence (Fig. 4). This residue is in a structurally conserved position to Ser-36 in DrMUG, and its side chain creates a barrier against the C5 position of bound uracil and prevents binding of a C5 methylated base. Mutating Tyr-147 in hUNG into Ala, Cys, or Ser gave an enzyme with the ability to excise thymine, although at a highly reduced rate compared with uracil excision by the wild-type enzyme. Furthermore, the hUNG Y147A mutant had a 10-fold higher thymine-excising activity than the Ser and Cys mutants, pinpointing the importance of this position in substrate specificity (30). Although human TDG has been found to mainly excise uracil and thymine from G:U and G:T mismatches, other TDGs have recently been identified as having other substrate specificities (30). Although human TDG has been suggested to be involved in substrate recognition, it is possible that the TDGs and MUGs having evolved different substrate recognition mechanisms. On the other hand, in the TDG orthologue from *S. pombe* (see above) the same residues are substituted into Gly and Ser (data not shown), similarly to ecMUG, which could elucidate the lack of affinity for G:T mismatches in these proteins. It thus appears that even within the TDG family of proteins different regions of the proteins may be involved in substrate interaction with the human TDG possibly having a more refined way of binding thymine in the substrate-binding pocket.

Prevention of binding of cytosine in the specificity pocket is probably aided by the side chain of Asp-84 (see above) in DrMUG (Fig. 4). In the crystal structure, the side chain of this residue is directed against the possible position of the amine-group of cytosine (Fig. 5). In ecMUG,

Further strengthening the possibility of the TDGs and MUGs having evolved different substrate recognition mechanisms. On the other hand, in the TDG orthologue from *S. pombe* (see above) the same residues are substituted into Gly and Ser (data not shown), similarly to ecMUG, which could elucidate the lack of affinity for G:T mismatches in these proteins. It thus appears that even within the TDG family of proteins different regions of the proteins may be involved in substrate interaction, with the human TDG possibly having a more refined way of binding thymine in the substrate-binding pocket.

Prevention of binding of cytosine in the specificity pocket is probably aided by the side chain of Asp-84 (see above) in DrMUG (Fig. 4). In the crystal structure, the side chain of this residue is directed against the possible position of the amine-group of cytosine (Fig. 5). In ecMUG,

Further strengthening the possibility of the TDGs and MUGs having evolved different substrate recognition mechanisms. On the other hand, in the TDG orthologue from *S. pombe* (see above) the same residues are substituted into Gly and Ser (data not shown), similarly to ecMUG, which could elucidate the lack of affinity for G:T mismatches in these proteins. It thus appears that even within the TDG family of proteins different regions of the proteins may be involved in substrate interaction, with the human TDG possibly having a more refined way of binding thymine in the substrate-binding pocket.

Prevention of binding of cytosine in the specificity pocket is probably aided by the side chain of Asp-84 (see above) in DrMUG (Fig. 4). In the crystal structure, the side chain of this residue is directed against the possible position of the amine-group of cytosine (Fig. 5). In ecMUG,

Further strengthening the possibility of the TDGs and MUGs having evolved different substrate recognition mechanisms. On the other hand, in the TDG orthologue from *S. pombe* (see above) the same residues are substituted into Gly and Ser (data not shown), similarly to ecMUG, which could elucidate the lack of affinity for G:T mismatches in these proteins. It thus appears that even within the TDG family of proteins different regions of the proteins may be involved in substrate interaction, with the human TDG possibly having a more refined way of binding thymine in the substrate-binding pocket.

Prevention of binding of cytosine in the specificity pocket is probably aided by the side chain of Asp-84 (see above) in DrMUG (Fig. 4). In the crystal structure, the side chain of this residue is directed against the possible position of the amine-group of cytosine (Fig. 5). In ecMUG,
Asp-84 is substituted by Lys-68, which has been suggested to provide some discrimination against cytosine and allow binding of 3-N\(^4\)-ethenocytosine (eC) (9). Binding of cytosine in the specificity pocket of hUNG is prevented by Asn-204. Mutation of this residue into Asp gave an enzyme with some ability to excise cytosine; however the affinity for uracil was still 100-fold higher than for cytosine (30). The reported affinity for cytosine of this hUNG mutant suggests that Asp-84 in drMUG may not fully prevent cytosine binding. There is an equivalent conservation of the residue in this position as an Asn in TDGs (Fig. 6), where it probably also participates in the substrate specificity by making specific contacts with the flipped-out DNA base in the bottom of the substrate-binding pocket. However, the TDGs also exhibit relatively high catalytic activity toward eC (29) and a firm conclusion with regard to substrate specificity should not be drawn, because no crystal structure of a TDG enzyme has been determined yet and the structural conservation between the MUG and TDG enzymes is not fully verified. The observed binding of acetate in the specificity pocket of drMUG further indicates that Asp-84 is protonated, i.e. in principle it is able to interact with both cytosine and uracil. However, the stability of the G:C pair in dsDNA will probably prevent drMUG from flipping a cytosine into the substrate-binding pocket and excising it. The possible cytotoxic consequences of excising cytosine from G:C base pairs would also rule against this catalytic function of drMUG. Further, characterization of drMUG showed that it had no catalytic activity against an oligonucleotide having a G:C pair (see below).

**Specificity**—ecMUG has been thoroughly studied with regard to its substrate specificity and was originally suggested to have a preference for G:U mismatches (9), although other studies have shown an even higher activity on G:T mismatches (29). When the damaged base is flipped into the substrate-binding pocket, it has been shown that Ser-145 and Arg-146 in ecMUG intercalate with the widened guanine on the complementary DNA strand, thus suggesting that damage recognition takes both DNA strands into account rather than only recognizing the mismatched base on the primary DNA strand (26). In drMUG, the suggested guanine-intercalating residues are substituted by Gly-164 and His-165, and an inspection of the structure does not suggest that any of these residues will intercalate with DNA bases. Based on this observation, drMUG seems to have guanine-independent substrate specificity (Fig. 7).

To investigate this hypothesis, the substrate specificity of both ecMUG and drMUG was studied in more detail using a gel shift assay. dsDNA oligonucleotides (1 pmol) with G:U, G:T, and A:U mismatches as well as a G:C match, in addition to ssU, were assayed as possible substrates for both enzymes (1 pmol). The results from this analysis show that ecMUG possesses specificity for the G:U-containing substrate only (Fig. 8), whereas drMUG has specificity for both G:U and A:U mismatches and ssU but not for G:T mismatches (Fig. 9). Thus, drMUG possesses a guanine-independent substrate specificity as predicted from the structural analysis.

In fact, the substrate specificity of drMUG is similar to the specificity identified for UNGs (10). However, because of the less well optimized catalytic residues of drMUG compared with UNGs (see below), we suspected that the specific activity would be lower than for UNG. The specific activity was therefore measured for drMUG using an A:U mismatch-containing substrate and compared with the specific activity of D. radiodurans UNG (drUNG) (33), an enzyme that possesses the conserved catalytic residues typical for UNGs (Fig. 4). The result shows that the specific activity of drMUG is only 10 units/mg, whereas it is 26,000 units/mg for drUNG (Table 1). Based on these results it appears that drUNG is the major A:U and G:U mismatch repair enzyme in D. radiodurans, whereas drMUG might serve as a backup system for UNG. This finding is in agreement with a recent study on uracil glycosylases from D. radiodurans (34). Another possibility is that drMUG serves as the main repair enzyme for other modified nucleotides that have not, to date, been identified. This may also be the case for ecMUG, which has been shown to possess high specificity toward several etheno adducts in DNA, e.g. eC (29), 5-hydroxymethyluracil (35), 1,N\(^4\)-ethenoguanine (36), and 8-(hydroxymethyl)-3,N\(^4\)-ethenocytosine (37). It is speculated that eC is the biologically relevant substrate for ecMUG because it is more active against this substrate than for uracil (38). The question of whether eC or other bases with specific damages caused by ionic radi-
...are the most relevant substrate for drMUG in nature is currently being investigated.

**Catalysis**—The catalytic residues Asp-145 and His-268 in hUNG have been found to be essential for catalysis, both by mutational analysis (39) and by computational studies (11). From the crystal structure of ecMUG it was suggested that Asn-18 (in the GINPG sequence), which is conserved in most known MUG/TDG sequences, binds a water molecule in a similar position and orientation to the catalytic aspartate of hUNG (Asp-145) and is aligned for nucleophilic attack on the deoxyribose in order for the catalysis to take place (Fig. 10A) (26). In drMUG these residues are substituted by Ala-34 and Ser-101 (Fig. 4); thus, drMUG will not be able to perform the equally efficient removal of uracil DNA as in hUNG. This is confirmed by the specific activity measurements performed in this work (Table 1).

The specificity analysis showed that drMUG was able to remove uracil from both GU and AU mismatches and ssU; hence, a catalytic residue in a different position might be involved in the catalysis. Structural analysis that involved modeling of drMUG with uracil in the substrate-binding pocket was therefore performed. Results from this analysis showed that the side chain of Asp-93 may participate in the catalysis indirectly via a nearby water molecule, which is also coordinated by the carbonyl oxygen of Gly-91 (Fig. 10B). To further investigate the involvement of Asp-93 in catalysis, a mutant drMUGD93A was produced and purified to apparent purity as shown in Fig. 1. The specificity analysis was performed as for the native protein, and the results from this experiment show that the mutant possesses no activity toward substrates containing GU or AU mismatches or ssU (Fig. 9) and also support the results from the structural analysis that suggested a central role for Asp-93 in drMUG-assisted catalysis.

To study the effect of the D93A substitution in more detail, the three-dimensional structure of the drMUGD93A mutant was determined to a resolution of 1.7 Å (Table 2). The refined electron density map clearly shows the substitution of residue 93 into Ala and also shows that the suggested catalytic water in drMUG has a slightly different position in drMUGD93A relative to the glycoside bond of the modeled uracil (Fig. 10C). The water molecule is still coordinated by the carbonyl oxygen of Gly-91; however, there are no other residues nearby that might fulfill the role of Asp-93 in order for a water molecule to perform an attack on the uracil substrate. Thus, the structure of drMUGD93A supports the results from activity measurements and shows that Asp-93 is essential for the excision of uracil from damaged DNA by drMUG.

**Evolution of MUG and TDG sequences among different species.** Phylogenetic tree (neighbor joining) of MUG and TDG sequences from *D. radiodurans* (deira; Q9RWF4), *B. japonicum* (braja; Q8JFB8), *R. palustris* (shopa; Q63N64), *Serratia marcescens* (serma; P43343), *S. thermoduricum* (synth; Q67QW8), *E. carotovora* (erevo; Q6D9D7), *S. coeliolor* (stco; Q9K3IS), *N. farrinica* (nafo; Q5YVW8), *G. gallus* (chicken; Q9P5UA), *M. musculus* (mouse; Q9Z1T3), *R. norvegicus* (rat; Q9N9NB), *H. sapiens* (human; Q13569), *X. laevis* (xenla; Q8AVR8), *X. tropicalis* (xentro; Q6DIK5), *E. coli* (ecoli; P33432), and *Salmonella typhimurium* (salty; Q7CR7). The codes in parentheses are the primary Swiss-Prot/TrEMBL accession numbers.

---

**Table 2**

| Data collection | drMUG | drMUGD93A |
|-----------------|-------|-----------|
| Resolution range (Å) | 12.1–1.75 (1.84–1.75) | 37.6–1.70 (1.79–1.70) |
| Number of unique reflections | 22,686 | 25,019 |
| Redundancy | 5.3 (5.1) | 7.0 (6.7) |
| Rmerge (%) | 5.4 (53.2) | 6.8 (59.6) |
| Completeness (%) | 99.9 (100) | 99.8 (100) |
| Mean I/σ(I) | 18.8 (2.8) | 18.3 (2.7) |

**Refinement statistics**

| R-factor (%) | 17.9 | 20.8 |
| Free R-factor (%) | 22.0 | 24.4 |
| Deviation from ideal geometry | 0.021 | 0.021 |
| Bond angles (°) | 1.916 | 1.877 |
| ESU (Å) (Rmerge/Rfree)**a** | 0.11/0.11 | 0.11/0.11 |
| Average | 24.3 | 22.9 |
| Side chain atoms | 27.4 | 25.2 |
| Acetate (number) | 24.9 (1) | 24.5 (1) |
| Water molecules (number) | 33.5 (151) | 31.7 (159) |
| All atoms | 26.5 | 24.8 |
| Ramachandran plot (%) | 92.4 | 89.6 |
| Most favored regions | 6.9 | 9.7 |
| Additional allowed regions | 0.7**a** | 0.0 |

---

**Figures**

**Figure 11.** Evolution of MUG and TDG sequences among different species. Phylogenetic tree (neighbor joining) of MUG and TDG sequences from *D. radiodurans* (deira; Q9RWF4), *B. japonicum* (braja; Q8JFB8), *R. palustris* (shopa; Q63N64), *Serratia marcescens* (serma; P43343), *S. thermoduricum* (synth; Q67QW8), *E. carotovora* (erevo; Q6D9D7), *S. coeliolor* (stco; Q9K3IS), *N. farrinica* (nafo; Q5YVW8), *G. gallus* (chicken; Q9P5UA), *M. musculus* (mouse; Q9Z1T3), *R. norvegicus* (rat; Q9N9NB), *H. sapiens* (human; Q13569), *X. laevis* (xenla; Q8AVR8), *X. tropicalis* (xentro; Q6DIK5), *E. coli* (ecoli; P33432), and *Salmonella typhimurium* (salty; Q7CR7). The codes in parentheses are the primary Swiss-Prot/TrEMBL accession numbers.
addition, Thr-121 and Ser-122 in the C-terminal part of β3 and Ala-126 in α4 are unique for the three enzymes. The conservation of Asp-93 among these sequences further supports its suggested role as a key residue for catalysis performed by drMUG.

The sequence alignment made by ClustalW was further processed with regard to phylogeny using MEGA2.1 (41). The results showed that drMUG, along with MUG from *B. japonicum* and *R. palustris*, clusters in one branch and seems to have evolved differently from eukaryotic TDG enzymes and other bacterial MUG enzymes (Fig. 11), suggesting that drMUG belongs to a new subfamily of the MUG/TDG glycosylases that we designate the MUG2 family. *B. japonicum* is an agriculturally important nitrogen-fixing symbiotic bacterium that is able to form root nodules on soybeans (42). *R. palustris* is a purple photosynthetic bacterium that belongs to the α-proteobacteria and is widely distributed in nature. It is a metabolically versatile bacterium that is able to use light, inorganic compounds, or organic compounds as energy sources (43).

**Concluding Remarks**—The crystal structure of the mismatch-specific uracil-DNA glycosylase from *D. radiodurans*, drMUG, is determined to a resolution of 1.75 Å. The structure of drMUG is overall the same as of *E. coli* MUG; however, specificity analysis shows that drMUG possesses a different substrate specificity compared with ecMUG and is able to remove uracil from both G:U and A:U mismatches in dsDNA as well as uracil in single-stranded DNA, whereas ecMUG only removes uracil from G:U mismatches. Analysis of sequence alignments and the crystal structure suggested that drMUG possesses a novel catalytic residue, Asp-93. Determination of the crystal structure of a mutant drMUGD93A to a resolution of 1.7 Å, and specificity analysis of this mutant confirmed that Asp-93 is crucial for catalysis. Finally, phylogenetic analysis suggested that drMUG is part of a new MUG/TDG subfamily (MUG2 family), together with *B. japonicum* and *R. palustris*, and that Asp-93 is conserved among these three enzymes. An obvious question is then why the enzymes in the MUG2 family possess a different catalytic residue compared with all other bacterial MUGs. We suggest that it has evolved as a response to need for broad substrate specificity in order to increase the DNA repair repertoire of these organisms. In the case of *D. radiodurans*, it probably has evolved in order to improve the DNA damage protection against radiation and desiccation.

A more in depth analysis of the substrate specificity is being initiated to possibly detect the most biological significant substrate of this enzyme. Mutational analysis is also in progress, the purpose of which is to get more insight into the catalytic mechanism of the MUG/TDG enzymes.