Structural Basis for Avoidance of Promutagenic DNA Repair by MutY Adenine DNA Glycosylase*

Lan Wang, Seung-Joo Lee, and Gregory L. Verdine

From the Departments of Chemistry and Chemical Biology, Stem Cell and Regenerative Biology, and Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138

Background: The mechanism by which MutY discriminates against the anti-substrate oxoG:C remains unclear.

Results: MutY binds the C of the anti-substrate oxoG:C in an exo-site.

Conclusion: MutY uses a combination of exo-site stabilization and active site destabilization to prevent pro-mutagenic processing of oxoG:C.

Significance: This work explains how MutY avoids accelerating mutagenesis by processing of an inappropriate substrate.

The highly mutagenic A:oxoG (8-oxoguanine) base pair in DNA most frequently arises by aberrant replication of the primary oxidative lesion C:oxoG. This lesion is particularly insidious because neither of its constituent nucleobases faithfully transmit genetic information from the original C:G base pair. Repair of A:oxoG is initiated by adenine DNA glycosylase, which catalyzes hydrolytic cleavage of the aberrant A nucleobase from the DNA backbone. These enzymes, MutY in bacteria and MUTYH in humans, scrupulously avoid processing of C:oxoG because cleavage of the C residue in C:oxoG would actually promote mutagenic conversion to A:oxoG. Here we analyze the structural basis for rejection of C:oxoG by MutY, using a synthetic crystallography approach to capture the enzyme in the process of inspecting the C:oxoG anti-substrate, with which it ordinarily binds only fleetingly. We find that MutY uses two distinct strategies to avoid presentation of C to the enzyme active site. Firstly, MutY possesses an exo-site that serves as a decoy for C, and secondly, repulsive forces with a key active site residue prevent stable insertion of C into the nucleobase recognition pocket within the enzyme active site.

The genomic integrity of living organisms is threatened by a wide variety of chemical and biological agents that damage DNA. One of the most common and genotoxic forms of damage is 8-oxoguanine (oxoG), which arises through the attack of reactive oxygen species on guanine (see Fig. 1A) (1). OxoG preferentially mispairs with adenine during replicative DNA polymerization, thereby giving rise to G:C to T:A transversion mutations (see Fig. 1B). Cellular resistance to the genotoxic effects of guanine oxidation is conferred by the evolutionarily conserved GO system, which consists of several proteins that intercept oxoG and its misreplication products at multiple stages of the mutagenesis process (2). The orthologous 8-oxoGTPases MutT (eubacteria) and hMTH1 (humans) sanitize the nucleotide precursor pool by directly hydrolyzing 8-oxoGTP to 8-oxo-dGMP; the paralogous DNA glycosylases MutM and hOGG1 (bacteria and humans, respectively) selectively cleave the oxoG nucleobase from oxoG:C base pairs (3, 4). Failure to repair oxoG:C results in the production of oxoG:A, a lesion that is particularly insidious because neither nucleobase conveys faithful information with which to direct repair of the other. Although the oxoG moiety is constitutionally aberrant and therefore readily distinguishable from undamaged DNA, the A moiety is only contextually aberrant and otherwise comprises one of four nucleobases in DNA; the difference between faithful repair and wreaking havoc on the genome therefore depends critically on recognition of the basepairing partner opposite A, being either oxoG (lesion) or T (undamaged). Repair of this lesion utilizes an oxoG:A-specific DNA glycosylase, bacterial MutY or human MUTYH, which leaves the oxoG unaffected and instead cleaves the contextually aberrant A, severing its glycosidic bond (see Fig. 1C) (3, 6). The resulting abasic site undergoes short-patch DNA repair resynthesis to furnish primarily oxoG:C, along with a lesser amount of oxoG:A. The latter is again acted upon by MutY/MUTYH, and the former is processed by MutM/hOGG1.

The severe risk posed by oxoG:A mispairs, and correspondingly, the criticality of oxoG:A repair by MutY and MUTYH, are underscored by the extraordinarily high rate of G:C to T:A transversion mutations in MutY-deficient bacteria (7, 8) and by the established role of MUTYH as a protector against genomic instability in humans. Loss-of-function mutations in MUTYH are associated with a high rate of G:C to T:A transversions in colorectal cancer patients (9). Furthermore, inactivation of MUTYH leads to tumorigenic mutations in other genes, including the tumor suppressor APC (9) and the protooncogene KRAS (10). The oxoG:A mispair can also be generated in DNA via replicative misincorporation of 8-oxoGTP; recent studies have shown that inhibition of hMTH1 elevates genomic
oxoG levels and is profoundly toxic to transformed human cells (11, 12). The realization that adenine DNA glycosylases are among the most important guardians of the genome has fueled interest in understanding the structure and function of MutY/MUTYH.

MutY and MUTYH belong to a large superfamily of structurally related DNA glycosylases having a signature helix-hairpin-helix motif followed by a Gly/Pro-rich loop and a key catalytic aspartic acid residue (HhH-GPD superfamily) (13, 14). The domain bearing the HhH-GPD motif lies at the N terminus of the protein and contains the catalytic pocket that recognizes and cleaves the substrate adenine (15, 16). MutY (and MUTYH) also contains a C-terminal domain that recognizes oxoG, thereby providing a means for discrimination of oxoG:A lesions versus undamaged T:A base pairs (17, 18). Equally important is avoidance of C excision from oxoG:C base pairs as this would accelerate transversion mutation by provoking repair-dependent formation of oxoG:A base pairs; indeed, MutY and mammalian MYH appear incapable of acting upon oxoG:C (19, 20). The structure of a lesion recognition complex having a catalytically incompetent but recognition-competent mutant MutY protein (D144N) bound to DNA bearing a central oxoG:A lesion has shed light on the mechanism of oxoG:A recognition (16). The substrate adenine is extruded from the DNA stack and inserted into an extrahelical active pocket, whereas the oxoG nucleobase is fully intrahelical, contacting primarily the N-terminal domain but also the C-terminal domain. In a related lesion recognition complex structure having a wild-type enzyme active site but an uncleavable “mutated” target nucleobase, 2′-fluoro-2′-deoxyadenosine, (FLRC, for fluorinated lesion recognition complex), the substrate adenine is plunged even more deeply into the active pocket, so as to reveal that adenine excision proceeds via general acid catalysis (21). The FLRC is therefore considered the structure that most accurately represents MutY poised to catalyze base excision. Although the molecular basis underlying recognition and excision of the oxoG:A substrate is well established both structurally and biochemically, the means by which MutY discriminates against oxoG:C as an anti-substrate remains a mystery. Here we have addressed this issue by analyzing at atomic resolution two complexes in which MutY has been captured in the act of interrogating a target oxoG:C base pair. These structures and the accompanying biochemical analysis suggest that MutY possesses a gatekeeper at the active site entrance that diverts C residues.

**Experimental Procedures**

*Geobacillus stearothermophilus* MutY Preparation—Point mutations (P164C, D144N) were introduced into the *G. stearothermophilus* MutY gene by QuikChange mutagenesis using a parent construct wherein the gene was cloned into the pET28 (Novagen) expression vector, resulting in an ORF with an N-terminal His tag and thrombin cleavage site. Protein was expressed and purified essentially as described before (16).

**Synthesis of Cross-linking DNA**—DNA oligomer 5′-TGGAC-3′ (X, modified A with a disulfide tether for cross-linking; O6-Phenyl-di-CE and oxoG phosphoramidite were purchased from Glen Research) was synthesized on an ABI 392 DNA synthesizer using standard reagents. To prepare the tether-containing DNA, the O6-Ph-dl-containing DNA was incubated with 1 ml of 1 m aqueous cysteamine containing 0.25 mM β-mercaptoethanol at 55 °C for 18 h. 1 ml of NH₂OH containing 0.25 mM β-mercaptoethanol was added to the resulting mixture and incubated at room temperature for 6 h to cleave off the DNA from the resin. The solution was desalted by NAP column chromatography (GE Healthcare), and concentrated by a SpeedVac. The residue was purified by urea PAGE, dissolved in 10 mM Tris, pH 8.0 buffer, and annealed with the complimentary strand (purchased from Eurofins MWG Operon).

**Glycosylase Assay of the Cross-linked Complex**—To test the cleavage activity of the cross-linked complex, we first cross-linked the oxoG-containing strand to the protein by mixing 40 μM protein and 27 μM DNA in buffer containing 15 mM Tris, pH 7.6, and 100 mM NaCl at 4 °C for 4 h. The resulting mixture was purified by MonoQ ion-exchange chromatography, buffer-exchanged to 20 mM Tris, pH 8, 200 mM NaCl, and then diluted to 4 μM. The 5′ end of the complimentary strand (containing either A or C at the position that base pairs with oxoG) was labeled with 32P and diluted to 200 nM with buffer containing 10 mM Tris, pH 8. 1 μl of the single-strand-cross-linked complex and 1 μl of the complimentary strand solution were added to 8 μl of reaction buffer containing 50 mM Tris, pH 7.4, 1 mM EDTA, 100 mM NaCl, resulting in a 10-μl reaction. At each time point recorded, 2 μl was taken out from the reaction mixture, and 5 μl of 2 M NaOH was added to quench the reaction. The mixture was heated up to 90 °C for 15 min to facilitate strand scission. An equal amount of formamide was added into the quenched reaction, and the sample was stored at −20 °C until subjected to gel analysis. Each sample was heated up to 90 °C for 4 min before being electrophoresed by 15% urea gel for 2 h at 35 mA. The gel was developed for 2 h and imaged with a PhosphorImager.

**Glycosylase Assay of the Catalytically Active MutY and the D144N Mutant MutY**—To test the cleavage activity of MutY in the uncross-linked condition, we prepared MutY in two forms: the catalytically active protein (bearing a P164C mutation for cross-linking and a wild-type active pocket) and the D144N MutY (also bearing a P164C mutation, as well as a D144N mutation in the active pocket). The 30-mer oligonucleotides (5′-TTTTTTTTTTGTCAXGTCCTTTTTTTTTTTT-3′; X is A in the oxoG:A DNA and C in the oxoG:C DNA) was end-labeled with 32P and annealed to the unlabelled complementary strand (5′-AAGAC®GTGGAC-3′; ® is 8-oxoguanine). The duplex was then incubated with catalytically active MutY or D144N MutY at room temperature for various amounts of time. Reactions were quenched by NaOH before they were analyzed by gel electrophoresis. The reaction conditions were similar to that used in the assay of the cross-linked complex.

**Protein-DNA Cross-Linking Reaction and Crystallization**—The protein-DNA complex was formed by incubating 40 μM protein and 27 μM DNA in buffer containing 15 mM Tris, pH 7.6, and 100 mM NaCl at 4 °C for 24 h. The complex was purified by MonoQ ion-exchange chromatography and dialyzed against buffer containing 10 mM Tris, pH 7.4, 90 mM NaCl, 10 μM β-mercaptoethanol at 4 °C overnight. The purified complex was concentrated to 175 μM and crystallized by the hanging-
Structure of MutY Bound to an Anti-substrate

drop vapor diffusion method at room temperature. The anti-substrate complex (ASC) was crystallized in 100 mM Tris, pH 8, 250 mM NaOAc, 29% (w/v) PEG 4000. The D144N ASC was crystallized in 100 mM Tris, pH 8.6, 1.75 M (NH₄)₂SO₄. Crystals appeared after several days and were allowed to grow for a couple of weeks. The ASC crystal was transferred to cryoprotectant solution containing 100 mM Tris, pH 8, 250 mM NaOAc, 29% (w/v) PEG 4000, and 10% glycerol; the D144N ASC crystal was transferred to cryoprotectant solution containing 100 mM Tris, pH 8.6, 1.75 M (NH₄)₂SO₄, and 25% glycerol. To reduce the ASC crystal, we transferred the crystal from the original reservoir solution to a reducing drop that contained 100 mM Tris, pH 8, 250 mM NaOAc, 29% (w/v) PEG 4000, and 25 mM tris(carboxyethyl)phosphine (TCEP). The drop was equilibrated against a reservoir that had the same composition under room temperature for 20 h. The resulting crystals were transferred to cryoprotectant solution containing 100 mM Tris, pH 8, 250 mM NaOAc, 29% (w/v) PEG 4000, 25 mM TCEP, and 10% glycerol and frozen into liquid nitrogen for data collection.

Data Collection and Structure Refinement—Data were collected on beamline 24-ID at the Northeastern Collaborative Access Team (NE-CAT) at the Advanced Photon Source at the Argonne National Laboratory and processed using XDS (22). Resolution cut-off was chosen by examining several indicators of data quality, such as I/σ(I), completeness, and CC1⁄2 of the highest resolution shell. Crystal of the ASC belongs to the space group P2₁2₁2₁, with unit cell dimensions of a = 38 Å, b = 85 Å, and c = 142 Å. Crystal of the D144N ASC belongs to the space group P6₁, with unit cell dimensions of a = 105 Å, b = 105 Å, and c = 235 Å. Data collection and refinement statistics are summarized in Table 1. The coordinates of the protein from the structure of P164C/D144N G. stearothermophilus MutY bound to oxoG:C-containing DNA (Protein Data Bank (PDB) ID: 1RRQ) were used as the initial search model. The model was improved by rigid body fit and subsequent simulated annealing, TLS (Translation/Libration/Screw), and individual B-factor refinement using Phenix (23). In both the ASC and the D144N ASC structures, the electron density of the extruded cytosine was clearly evident in their respective positions in the αA-weighted Fo − Fc map after molecular replacement. The model was improved by rigid body fit and subsequent simulated annealing, TLS (Translation/Libration/Screw), and individual B-factor refinement using Phenix (23). In both the ASC and the D144N ASC structures, the electron density of the extruded cytosine was clearly evident in their respective positions in the αA-weighted Fo − Fc map after the adjacent nucleotides were put into the model. Model building was performed in Coot (24) while constantly monitoring Rfree (25). Water molecules were added to the model using both automated methods in Phenix (23) and manual inspection of difference maps. Figures were prepared using PyMOL version 1.7 (26).

Results

Experimental System—Previous work has demonstrated that MutY does not readily crystallize with duplex DNA containing a target oxoG:A base pair, but that crystallization of this complex can be fostered through introduction of an intermolecular disulfide cross-link at a site several base pairs removed from the target base pair (Fig. 1, D and E) (16, 27). Anticipating the necessity to provide similarly for stabilization of MutY on an oxoG:C target base pair from which it would ordinarily dissociate, we again employed disulfide cross-linking in the studies described below. The previously established cross-linking system was used here (16); the protein component is MutY wild type apart from having an engineered Cys residue at position 164, and the DNA component comprises two complementary 11-mer paired to leave a one-nucleotide overhang at each end and with a C2 thiol attached to nucleobase A4 in the major groove. The resulting cross-linked complex furnished, upon optimization, high-quality crystals that enabled structure determination of the ASC via molecular replacement and model building with refinement to 2.2 Å resolution (Table 1).

General Features of MutY Interrogating an Anti-substrate oxoG:C Base Pair—The overall architecture of the ASC is very similar to that of the FLRC (heavy atom root mean square deviation = 0.323 Å; Fig. 2, A and B), with the protein completely encircling a DNA duplex sharply bent at the site of the target oxoG:C base pair. The N-terminal domain penetrates the duplex from the minor groove side opposite the fully intrahelical oxoG nucleobase, engaging the Watson-Crick edge of the oxoG in hydrogen-bonding interactions, whereas the C-terminal domain engages oxoG from the major groove side (Figs. 2B and 3H). Indeed, nearly identical in the ASC and FLRC structures are the protein component plus the entire oxoG-containing strand and even the flanking regions of the strand containing the nucleobase targeted for excision. Notwithstanding these similarities, pronounced structural divergence is apparent in the immediate vicinity of the target nucleobase, A (FLRC) or C (ASC) (compare Fig. 2, A and C versus B and D, respectively; see also Fig. 3, A–D). Although the target nucleobase is fully extrahelical in both the FLRC and the ASC structures, the actual location of the target nucleobase within the complex and the local conformation of the DNA backbone at that site bear marked divergence. In FLRC, the substrate mimetic 2'-β-fluoro-adenosine is deeply inserted into the active pocket, which lies at the junction between the two structural modules that comprise the catalytic domain (the six-helix barrel module and the [4Fe-4S] cluster module). By contrast, in the ASC, the anti-substrate C is not plunged into the active site pocket but instead is nestled into an exo-site groove formed by a helix-hairpin-helix motif (α8, α9, and the loop in between) belonging to the [4Fe-4S] module (Fig. 2, C and D).

Although MutY bears a fully wild-type active site, the structure reveals continuous electron density throughout the 2'-deoxyribose and C nucleobase moieties, indicative of an intact glycosidic bond (Fig. 3E). Disulfide cross-linking has occasionally been observed to enable cleavage of an ordinarily dispreferred substrate, and the lack of any apparent cleavage in the ASC suggested that the MutY active site possesses extremely strong discrimination against the anti-substrate C. To rule out the possibility that disulfide cross-linking per se abrogates the catalytic activity of MutY, we turned to in vitro cleavage assays on disulfide cross-linked MutY complexes having either a substrate A or an anti-substrate C at the site of the target nucleobase. The reaction was initiated by adding the (anti) substrate-containing strand to a complex of MutY disulfide cross-linked to the oxoG-containing strand. As shown in Fig. 4, A and B, the

4 M. Zhou and G. Verdine, unpublished data.
FIGURE 1. Generation and repair of oxoG:C and oxoG:A lesions in DNA. A, the primary oxoG lesion is formed by attack of reactive oxygen species (ROS) on a guanine residue. B, pathway for repair of oxoG lesions in DNA. C, hydrolytic cleavage of the target glycosidic bond catalyzed by MutY. D, duplex DNA sequence used in this work. The arrow indicates the adenine that is modified (as shown in panel E) to cross-link to an engineered Cys residue in MutY. E, structure of the modified adenine used in disulfide cross-linking and of the covalent linkage it forms with MutY.
FIGURE 2. Overall structures of the ASC and the FLRC. A, structure of the FLRC (PDB ID: 3G0Q) showing 2'-β-fluoro-oxoadenosine (FA) in the active site. The N-terminal catalytic domain of MutY is colored in cyan, the C-terminal MutT domain is in blue, and the iron-sulfur cluster is shown as orange (iron) and yellow (sulfur) sticks. The DNA backbone is colored green, and the oxoG and 2'-β-fluoro-oxoadenosine are shown as red (oxoG) and orange (A) spheres.

B, structure of the ASC (present work) showing C in the exo-site. The color scheme is the same as in A except that DNA backbone is colored gold and C is in magenta. C and D, the bottom panel shows the FLRC (C) and the ASC (D) structures rotated 90 degrees from the top panel view, with the protein shown in surface representation.

TABLE 1

Data collection and refinement statistics

|                       | ASC             | Reduced ASC     | D144N ASC        |
|-----------------------|-----------------|-----------------|------------------|
| Resolution range (Å)  | 41.41–2.21 (2.29–2.21) | 41.37–2.32 (2.40–2.32) | 85.01–2.59 (2.68–2.59) |
| Space group           | P21212          | P212121         | P61              |
| a, Å                  | 37.7            | 37.6            | 105.3            |
| b, Å                  | 84.6            | 83.8            | 105.3            |
| c, Å                  | 142.5           | 142.7           | 235.4            |
| α, °                  | 90              | 90              | 90               |
| β, °                  | 90              | 90              | 90               |
| γ, °                  | 90              | 90              | 120              |
| Total reflections     | 166,023 (15,823) | 73,154 (73,44)  | 264,923 (26,696) |
| Unique reflections    | 23,611 (22,86)  | 20,028 (19,92)  | 45,769 (46,10)   |
| Completeness (%)      | 99.86 (99.87)   | 98.45 (100.00)  | 99.97 (99.85)    |
| Multiplicity (%)      | 7.0 (6.9)       | 3.6 (3.7)       | 5.8 (5.8)        |
| Mean I/σ(I)          | 12.67 (1.31)    | 9.11 (1.45)     | 15.60 (1.66)     |
| Rmerge, a,b           | 0.1275 (1.477)  | 0.118 (0.8846)  | 0.1212 (1.209)   |

Refinement statistics

|                       | ASC             | Reduced ASC     | D144N ASC        |
|-----------------------|-----------------|-----------------|------------------|
| Rwork (%)             | 18.44 (30.68)   | 18.70 (26.78)   | 16.94 (28.38)    |
| Rfree (%)             | 24.96 (38.85)   | 25.89 (30.43)   | 22.80 (32.22)    |
| Number of non-hydrogen atoms | 3333            | 3260            | 6575             |
| Macromolecules        | 3191            | 3191            | 6530             |
| Ligands               | 15              | 9               | 16               |
| Water                 | 127             | 60              | 29               |
| Protein residues      | 366             | 366             | 749              |
| r.m.s. deviation from ideality | 0.01            | 0.011           | 0.012            |
| Bond length (Å)       | 1.4             | 1.37            | 1.45             |
| Ramachandran plot (%) | 96.2            | 95.6            | 92.1             |
| Favored               | 3.2             | 3.5             | 5.3              |
| Allowed               | 0.6             | 0.9             | 2.6              |
| Average B-factor      | 57.6            | 64.5            | 57.7             |
| Macromolecules        | 58              | 64.8            | 57.8             |
| Ligands               | 54.7            | 46.7            | 38.2             |
| Solvent               | 48.3            | 53.7            | 44.6             |

PDB ID

|               | ASC             | Reduced ASC     | D144N ASC        |
|---------------|-----------------|-----------------|------------------|
| 4YOQ         | 4YPH            | 4YPR            |                  |

* Values in parenthesis refer to the highest resolution shell.

b $R_{merge} = \sqrt{\sum I - \langle I \rangle}/\sqrt{\sum I}$; where $I$ is the observed intensity.

c $R_{work} = \sqrt{\sum |F_0| - |F_c|}/\sum |F_0|$; where $F_0$ and $F_c$ are the observed and calculated structure factor amplitudes, respectively.

d $R_{free}$ was calculated based on 5% data randomly selected and omitted throughout structure refinement (32).

e r.m.s., root mean square.

f Calculated using MOLPROBITY (33).
substrate-containing strand underwent rapid cleavage on the time-scale of minutes (60% within 10 min), whereas the anti-substrate-containing strand was intact after 2 days. Thus, the powerful discrimination of A versus C is an intrinsic property of MutY and is observed even in the extreme situation in which the enzyme and DNA are physically linked; these results reinforce the notion of C as a true anti-substrate for MutY.

Factors That Determine the DNA Conformation in the ASC—As mentioned above, the target strand DNA conformation of the ASC diverges greatly from that in the FLRC. Although the FLRC shows the DNA backbone at the target site hyperextended out over the minor groove so as to enable A to reach the distal active site pocket, in the ASC, the backbone lies over the major groove and is considerably less extended from the duplex (Fig. 3A). This divergence of backbone conformation is localized nearly completely to the site of the target nucleobase, with the difference in the two complexes being 1.9 Å at p1, 11.2 Å at p0, and 0.6 Å at p−1 (where p indicates phosphate). Consistent with this profound but highly localized divergence, the amino acid contacts to p−1 and p1 are conserved in the FLRC and
Structure of MutY Bound to an Anti-substrate

To account for the influence of the disulfide cross-link on the structure of the complex derived thereby, we reduced treated crystal of the complex with a powerful disulfide-reducing agent, TCEP. Previous attempts in other systems to employ this strategy resulted in a loss of crystal integrity, but in the present case, no such loss was seen, and therefore we solved the structure of the complex produced by TCEP reduction to 2.3 Å resolution (Table 1). The resulting structure clearly exhibited a disappearance of the positive density seen in the ASC between the attachment points for the cross-link, Cys-164 and A-4. The local segment of protein structure in the vicinity of Cys-164 also differs slightly in the disulfide cross-linked and reduced complexes. Together, the loss of density for the disulfide, the adjustment of the loop bearing Cys-164, and the clear electron density for the loop in the TCEP-treated complex indicate that the disulfide cross-link had indeed undergone reductive cleavage to a dithiol (Fig. 5, A–C). In Fz – Fs maps of the reduced complex prior to modeling of the target cytosine, positive electron density was observed at the same exo-site position in which the extruded cytosine base had been modeled in the disulfide cross-linked complex, thereby confirming that the target cytosine remained bound in the exo-site (Fig. 5E). In both the non-reduced and the reduced complexes, the main-chain carbonyls of the cross-linked residue, Cys-164, and its N-terminal neighbor, Lys-163, appear to hydrogen-bond with the N4-H of the extruded cytosine base, although the contact distance and geometry of the Cys-164 are rendered less favorable in the reduced complex because of the loop movement described above. Given that

\* L. Wang and G. Verdine, unpublished data.

FIGURE 4. Glycosylase assay of MutY. A, a representative urea gel showing cleavage of A opposite oxoG by MutY (left set of lanes) and lack of cleavage of C opposite oxoG under the same conditions. (right set of lanes) To test the glycosylase activity of the cross-linked complex used in crystallography, we cross-linked the strand containing oxoG to the protein and then added to it the 32P-labeled complimentary strand (containing either A or C to pair with oxoG). The reaction was quenched at various time points by adding NaOH into the mixture and heating at 90 °C for 15 min (see “Experimental Procedures” for detailed procedure). B, fraction of DNA cleaved in the two reactions in A plotted as a function of time. Each data point shows an average of three parallel experiments, and the error bar shows the standard deviation from the average. C, cleavage assays of catalytically active MutY (containing only the P164C mutation used in disulfide cross-linking) and the D144N mutant protein (also containing the P164C mutation) (neither of the two proteins is cross-linked to DNA in this experiment). During the course of the experiment, the active P164C MutY achieved almost 100% cleavage of the oxoG:A base pair, whereas no cleavage product was detected for the D144N mutant protein. Neither the catalytically active MutY nor the D144N mutant protein showed any detectable activity toward the anti-substrate oxoG:C.
residue 164 appears to play such an important role in stabilizing the extrahelical Cys in the exo-site, we sought to account for the effect of mutating residue-164 from Pro to Cys. Thus, we superimposed the wild-type protein from the (uncross-linked) abasic site complex (PDB ID: 1RRS) with the protein in the ASC (having Pro-164 and Cys-164, respectively) and found that with the wild-type protein, the exo-site cytosine remains within hydrogen-bonding distance of the backbone carbonyls of Pro-164 and Lys-163 (Fig. 5, G and H). This analysis clearly indicates that wild-type MutY bears the same exo-site as P164C MutY and hence is capable of binding an extrahelical cytosine in analogous fashion to that seen in the ASC.

The Active Site D144N Mutation in the ASC (D144N ASC) Induces Translocation of the Anti-substrate C from the Exo-site to the Active Site—The active site residue Asp-144, which is known to be critical for catalysis by MutY (15, 28) (see also Fig. 4C), is believed to stabilize the oxocarbenium intermediate in adenine excision. Asp-144 also plays a crucial role in enabling full insertion and correct positioning of the substrate adenine within the active site (21); specifically, mutation of Asp-144 to Asn yielded a lesion recognition complex in which the substrate adenine is partially dislodged from the active site pocket. To assess the impact of the D144N mutation on the structure of the ASC, we cross-linked the D144N protein to DNA in the same fashion as described above and solved the structure of this complex (D144N ASC) to 2.6 Å resolution (Table 1). The asymmetric unit comprises two nearly identical complexes in the crystal structure of the D144N ASC (heavy atom root mean square deviation 0.275 Å); we chose protomer B for our analysis. The D144N mutation has little effect on the disposition of the protein and most of the DNA duplex, as judged by comparison of the ASC structure with that of D144N ASC (Fig. 6A). Remarkably however, the D144N mutation alone is sufficient to cause the anti-substrate C to transition from the exo-site to the active site pocket, via a highly localized but substantial change in the conformation of the target DNA strand at the locus of the anti-substrate C (Fig. 6A). The residue corresponding to Asp-144 bears the same exo-site as P164C MutY and hence is capable of binding an extrahelical cytosine in analogous fashion to that seen in the ASC.

FIGURE 5. Structure of the reduced ASC. A, close-up view of the cross-linking site in the ASC, colored as in Fig. 3C. B, close-up view of the cross-linking site in the reduced ASC. Protein and DNA are in pink. C, protein superposition of panels A and B. D, exo-site in the ASC, colored as in Fig. 2B. Important hydrogen-bonding interactions between the protein and the cytosine are denoted in dotted lines, with the associated numbers, showing distances in Å. E, exo-site in the reduced ASC, colored as in panel B. F, protein superposition of panels D and E. G, superposition of the wild-type protein in the abasic site complex (PDB ID: 1RRS) with the protein in the ASC. The wild-type protein (with Pro-164) and the DNA in ASC are shown. H, superposition of panels D and G. The αA-weighted 2F0 − Fc map is contoured at 1σ in panels A, B, and E. The Fc − Fσ map is contoured at 3σ in panels A and B and at 2.5σ in panel E.
question (Asp-144 carboxylate-O^- and O4^-) are only 3.6 Å apart, close enough that the two probably experience a repulsive interaction in the ground state; this unfavorable interaction would also drive catalysis via bound state destabilization (30, 31). This putative repulsive interaction would explain the conformational effect of the D144N mutation as the change from a carboxylate to carbamamide side chain at position 144 would replace a repulsive interaction with an attractive one. This same train of logic suggests that the Asp-144/O4^- interaction is repulsive enough to disfavor stable binding of the anti-substrate C to the enzyme active site.

Whether the target nucleoside prefers to bind in the exo-site or active site is determined by the balance of forces originating not only from the attractive versus repulsive interactions between Asp-144 and the substrate sugar moiety, but also the interactions established between the target nucleobase and its binding pocket within the enzyme active site. This nucleobase-specific interaction repertoire provides an explanation as to why C in the anti-substrate complex prefers an exo-site disposition, whereas A in the physiologic substrate prefers the active site pocket. Finally, the observation that even when forced to remain bound to each other for 2 days, the anti-substrate fails to undergo detectable cleavage by MutY suggests the likelihood that there exists yet a third proofreading mechanism, namely the inability of the enzyme to catalyze cleavage of C should it perchance enter the active site.

**Discussion**

To perform faithfully their role as guardians of the genome, DNA glycosylases must not only cleave their cognate substrate, but must also avoid cleaving or even binding stably to non-substrates. The stakes are particularly high with MutY and C:oxoG as cleavage of the C in this base pair would provoke error-prone repair replication and hence accelerate the mutagenic process. For this reason, we refer to C:oxoG as an anti-substrate for all adenine DNA glycosylases. The results presented here provide a structural rationale for the mechanism by which MutY avoids pro-mutagenic repair of C:oxoG. We propose that MutY uses at least two distinct mechanisms to prevent accidental cleavage of a bound C:oxoG that has undergone extrusion of the target C from the helical stack. Firstly, the exo-site serves as a decoy that stabilizes the extrahelical cytosine at a location remote from the enzyme active site. Secondly, even if C manages to shake off the grasp of the exo-site, the enzyme active site rejects the non-cognate nucleoside because the balance of attractive and repulsive forces between the 2'-deoxyribose moiety and Asp-144, and between C and the active site pocket, disfavors engagement of C in the active site pocket. Finally, the observation that even when forced to remain bound to each other for 2 days, the anti-substrate fails to undergo detectable cleavage by MutY suggests the likelihood that there exists yet a third proofreading mechanism, namely the inability of the enzyme to catalyze cleavage of C should it perchance enter the active site.

**Author Contributions**—L.W. and G.V. designed the study and wrote the manuscript. L.W. performed the experiments. All authors analyzed the results and approved the final version of the manuscript.

**Acknowledgments**—We thank the entire staff of beamline 24-ID of the Advance Photon Source at Argonne National Laboratory for expert assistance with x-ray data collection and processing.
References

1. Lindahl, T. (1993) Instability and decay of the primary structure of DNA. Nature 362, 709–715
2. David, S. S., O’Shea, V. L., and Kundu, S. (2007) Base-excision repair of oxidative DNA damage. Nature 447, 941–950
3. Michaels, M. L., Cruz, C., Grollman, A. P., and Miller, J. H. (1992) Evidence that MutY and MutM combine to prevent mutations by an oxidatively damaged form of guanine in DNA. Proc. Natl. Acad. Sci. U.S.A. 89, 7022–7025
4. Michaels, M. L., and Miller, J. H. (1992) The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-oxoguanine. J. Bacteriol. 174, 6321–6325
5. Shibuuti, S., Takeda, M., and Grollman, A. P. (1991) Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. Nature 349, 431–434
6. Slupska, M. M., Baikalov, C., Luther, W. M., Chiang, J. H., Wei, Y. F., and Miller, J. H. (1996) Cloning and sequencing a human homolog (hMYH) of the Escherichia coli mutY gene whose function is required for the repair of oxidative DNA damage. J. Bacteriol. 178, 3885–3892
7. Moriya, M., and Grollman, A. P. (1993) Mutations in the mutY gene of Escherichia coli enhance the frequency of targeted G:C to T:A transversions induced by a single 8-oxoguanine residue in single-stranded DNA. Nucleic Acids Res. 21, 431–434
8. Ngiem, Y., Cabrera, M., Cupples, C. G., and Miller, J. H. (1988) The mutY gene: a mutator locus in Escherichia coli that generates G:C to T:A transversions. Proc. Natl. Acad. Sci. U.S.A. 85, 2709–2713
9. Al-Tassan, N., Chielig, N. H., Maynard, J., Flemming, N., Livingston, A. L., Williams, G. T., Hodges, A. K., Davies, D. R., David, S. S., Sampson, J. R., and Cheadle, J. P. (2002) Inherited variants of MYH associated with sporadic G:C to T:A mutations in colorectal tumors. Nat. Genet. 30, 227–232
10. Lippton, L., Halford, S. E., Johnson, V., Novelli, M. R., Jones, A., Cummings, C., Barclay, E., Sieber, O., Sadat, A., Bisgaard, M.-L., Hodgson, S. V., Aaltonen, L. A., Thomas, H. J. W., and Tomlinson, I. P. M. (2003) Carcinogenesis in MYH-associated polyps follows a distinct genetic pathway. Cancer Res. 63, 7595–7599
11. Gad, H., Koolmeister, T., Jenhm, A.-S., Eshtad, S., Jacobs, S. A., Ström, C. E., Svensson, L. M., Schultz, N., Lundbäck, T., Einarsdóttir, B. O., Saleh, A., Göktürk, C., Baranczewski, P., Svensson, R., Berntsson, R. P.-A., Gustafsson, R., Strömberg, K., Sanjiv, K., Jacques-Cordonnier, M.-C., Dessertes, M., Gustavsson, A.-L., Olofsson, R., Johansson, F., Homan, E. J., Loseva, O., Bräutigam, L., Johannsson, L., Höglund, A., Hagenkort, A., Pham, T., Altman, K., Gaugaz, F. Z., Vikking, S., Evers, B., Henrikson, M., Völz, K. S. A., Wallner, O. A., Hammarsström, L. G. J., Wiita, E., Almöst, I., Kalderón, C., Axelsson, H., Djureinovic, T., Puisyvert, J. C., Hägglad, M., Jeppson, P., Martens, U., Lundin, C., Lundgren, B., Granelli, L., Jensen, A. S., Artursson, P., Nilsson, J. A., Stenmark, P., Scobie, M., Berglund, U. W., and Hellday, T. (2014) MTH1 inhibition eradicates cancer by preventing sanitation of the DNTP pool. Nature 508, 215–221
12. Huber, K. V. M., Salah, E., Radic, B., Gridling, M., Elkins, J. M., Stukalov, A., Jenhm, A.-S., Göktürk, C., Tariq, K., Strömberg, K., Pham, T., Berglund, U. W., Colinge, J., Bennett, K. L., Loizou, J. I., Hellday, T., Knapp, S., and Superti-Furga, G. (2014) Stereoscopic targeting of MTH1 by (S)-crizotinib as an anticancer strategy. Nature 508, 222–227
13. Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P., and Tainer, J. A. (1995) Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. EMBO J. 14, 4108–4120
14. Nash, H. M., Bruner, S. D., Schärer, O. D., Kawate, T., Addona, T. A., Spooner, E., Lane, W. S., and Verdine, G. L. (1996) Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA repair protein superfamily. Curr. Biol. 6, 968–980