Entrapment and Structure of an Extrahelical Guanine Attempting to Enter the Active Site of a Bacterial DNA Glycosylase, MutM*§

Yan Qi†, Marie C. Spong‡, Kwangho Nam§, Martin Karplus¶, and Gregory L. Verdine**

From the †Graduate Program in Biophysics, Harvard Medical School, Boston, Massachusetts 02115, the Departments of ‡Chemistry and Chemical Biology, §Molecular and Cellular Biology, and ¶Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, the **Chemical Biology Initiative and Program in Cancer Chemical Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, and the §Institut de Science et d’Ingénierie Supramoléculaires, Université Louis Pasteur, 67000 Strasbourg, France

MutM, a bacterial DNA glycosylase, protects genome integrity by catalyzing glycosidic bond cleavage of 8-oxoguanine (oxoG) lesions, thereby initiating base excision DNA repair. The process of searching for and locating oxoG lesions is especially challenging, because of the close structural resemblance of oxoG to its million-fold more abundant progenitor, G. Extrusion of the target nucleobase from the DNA double helix to an extrahelical position is an essential step in lesion recognition and catalysis by MutM. Although the interactions between the extruded oxoG and the active site of MutM have been well characterized, little is known in structural detail regarding the interrogation of extruded normal DNA bases by MutM. Here we report the capture and structural elucidation of a complex in which MutM is attempting to present an undamaged G to its active site. The structure of this MutM-extrahelical G complex provides insights into the mechanism MutM employs to discriminate against extrahelical normal DNA bases and into the base excision process in general.

Cells are constantly challenged by both endogenous and exogenous sources of DNA damage, such as reactive oxygen species generated during metabolism and chemicals from the environment (1). Failure to repair the resulting damaged nucleobases has adverse consequences ranging from acquisition of mutations to induction of cell cycle arrest and apoptosis (2). Cells have evolved mechanisms that protect genome integrity by specifically identifying and repairing lesions within their DNA. The base excision DNA repair pathway, for example, is responsible for the recognition and repair of single-nucleobase lesions. Base excision DNA repair is initiated by DNA glycosylases, which target aberrant nucleobases and catalyze cleavage of their glycosidic linkage to the DNA backbone (3). Despite the structural divergence of the target lesions for these enzymes, all of the known DNA glycosylases gain access to their target nucleobases by extruding the entire target nucleoside from the DNA helix and inserting it into an extrahelical active site pocket (4).

MutM (also known as Fpg) is a well characterized bacterial DNA glycosylase that specifically recognizes and cleaves oxidatively damaged nucleobases (5). The major substrate, oxoG, is the predominant form of oxidative damage in DNA, because G has the highest redox potential among all four normal nucleobases (6). During DNA replication, adenine (A) is frequently incorporated opposite oxoG through Hoogsteen base pairing; replication of this intermediate results in a G:C to T:A transversion mutation (7). Deficiency in the human functional counterpart of MutM, hOgg1, has been implicated in a variety of cancers (8).

MutM faces the formidable challenge of locating rare oxoG sites amid a vast excess of undamaged DNA, and doing so before replication unleashes the mutagenic potential of the lesion. In mammalian cells, oxoG occurs at a steady state frequency of $10^{-7}–10^{-8}$/nucleotide, more than a million-fold less frequent than G (9). Unlike many other types of DNA damage (thymine dimer, ring-opened bases, and alkylation) that impose characteristic perturbations to the local DNA conformation, oxoG has minimal impact on the structure and energetics of B-form DNA (10–12). The absence of structural alterations in isolated oxoG-containing DNA duplex is not surprising because the Watson-Crick face is unchanged, and oxoG differs from G by only two atoms: an additional oxygen present on the C-8 position of oxoG and a lone pair of electrons on N-7 of G in place of an N-7-hydrogen in oxoG (Fig. 1A). Further complicating the search process is the fact that a necessary step in repair by MutM is complete extrusion of the target nucleoside from the DNA helical stack and insertion into the extrahelical active

---

* This work was supported, in whole or in part, by National Institutes of Health Grants GM044853, GM047467, and CA100742 (to G. L. V.) and GM030804 (to M. K.).
† The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1–S8.
The atomic coordinates and structure factors (codes 3JR4 and 3JR5) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).
§ Supported by a predoctoral fellowship from the National Science Foundation.
¶ Supported by a predoctoral fellowship from the Howard Hughes Medical Institute.
* Supported by a postdoctoral fellowship from the National Cancer Center.
** To whom correspondence should be addressed: 12 Oxford St., Cambridge, MA 02138. Fax: 617-495-8755; E-mail: gregory_verdine@harvard.edu.

The abbreviations used are: oxoG, 8-oxoguanine; DXL, disulfide cross-linking; OCL, oxoG-capping loop; LRC, lesion recognition complex; XGC, extrahelical G complex; IC, interrogation complex; EC, encounter complex; WT, wild type; MD, molecular dynamics.
How does MutM locate oxoG with such high precision and efficiency? Recent structural studies of MutM-DNA complexes at various stages along the base extrusion pathway have provided valuable insights. As shown by the structural studies and computations (14), the extrusion pathway consists of (i) initial disengagement of the target nucleobase from its complementary pairing partner and transition to an unstacked and exposed conformation, followed by (ii) rotation of the extrahelical base about its glycosidic linkage to a syn configuration, and then (iii) swiveling of the DNA backbone to enable displacement of the target nucleoside away from the DNA surface and insertion of the target base in the enzyme active site, accompanied by structural adaptations in the enzyme. Detailed views of the end state of the base extrusion pathway (Fig. 1B) have been gained via structures of lesion recognition complexes (LRCs) comprising a catalytically inactive but recognition-competent mutant of MutM bound to DNA with an extrahelical oxoG in the active site. These structures revealed extensive contacts to the Hoogsteen face of the extruded oxoG by a loop on the enzyme, the oxoG-capping loop (OCL), that becomes ordered only upon full insertion of oxoG into the enzyme active site (15). Rational design of a disulfide cross-linking (DXL) system based on the LRC structures enabled entrapment and structural elucidation of MutM interrogating intrahelical nucleobases, either normal or damaged, at the earliest stages of the base extrusion pathway. Specifically, the structures of “interrogation complexes” (ICs) having MutM interrogating a fully base-paired undamaged target nucleobase suggested that MutM actively interrogates the intact DNA duplex while searching for lesions (16). By destabilizing the interaction between oxoG and the OCL, we have been able to trap “encounter complexes” (ECs), in which MutM is examining a fully intrahelical target oxoG (14). Comparison of several pairs of sequence-matched EC and IC structures, the protein-DNA interfaces of which differ only by the two atoms that distinguish oxoG from G, illuminated the structural basis of the ability of MutM to discriminate an intrahelical lesion from its normal progenitor. This structural difference underlies the kinetic preference of MutM to extrude oxoG from DNA more rapidly than G. Notwithstanding these advances in elucidation of the MutM base extrusion pathway, to this date, no complex of MutM interrogating an extrahelical G has been structurally characterized, because this state of the system is ordinarily too fleeting to be observed crystallographically.

Here, we report the entrapment and structural elucidation of a MutM-DNA complex trapped at the stage of attempting to present an extrahelical, undamaged G to the active site. As described in detail below, the entrapment of this intermediate was enabled by the use of DXL technology to tug on the DNA backbone at the site of the target nucleobase. The structure of this MutM-extrahelical G complex (XGC) reveals that the MutM active site can detect the subtle structural differences between extrahelical G and oxoG, so that an extruded normal G fails to attain the final precatalytic state of the base extrusion pathway.

EXPERIMENTAL PROCEDURES

Cloning, Overexpression, and Purification of MutM—The N174C point mutation was introduced into WT or the enzymatically inactive E3Q mutant of Geobacillus stearothermophilus MutM (15) in a pET24 (Novagen) expression vector using a QuikChange II site-directed mutagenesis kit (Stratagene). N174C MutM and N174C E3Q MutM were overexpressed and purified essentially as described before (17).

DNA Synthesis and Purification—DNA oligomers containing the backbone DXL N-ethylthio tethers were synthesized on an ABI 392 DNA synthesizer (Applied Biosystems) using published methods (16). The complementary strands of normal DNA oligomers were synthesized on a Mermade-12 Oligonucleotide Synthesizer (BioAutomation) using standard phosphoramidite chemistry. DNA oligomers were purified by urea-PAGE, dissolved in 10 mm Tris, pH 8.0, and annealed at 200 μM concentration. Sequence and disulfide functionalization of the oligomers were confirmed by matrix-assisted laser desorption
Structure of MutM with an Extrahelical Guanine

ionization. The following DNA oligomer duplexes were prepared for the structures discussed in this paper, with “G denoting oxoG and x representing the location of the phosphate moiety bearing the nonbridging N-ethylthio-tether modified phosphoramide used in cross-linking to MutM: lesion recognition complex (LRCN174C), 5'-TGCGTCCTGCAGTGGAG-3'; XGC, 5'-TGCGTCACTCC-3' and 3'-CGCAGGTCCAGATGGA-5'.

DXL Reactions and Crystallization—LRCN174C and XGC were made by cross-linking the corresponding DNA duplexes to the N174C E3Q MutM protein. Preparative cross-linking reactions were performed under nondenaturing conditions as previously described (16). The DXLed complexes were purified on a MonoQ column using a linear gradient of 100–600 mM NaCl in 20 mM Tris, pH 7.4, buffer exchanged into 20 mM Tris, pH 7.4, and 50 mM NaCl, concentrated to 250 μM, and subjected to crystallization under essentially the same conditions as before (16). Crystals readily appeared within a week at 4 °C using the hanging drop vapor diffusion method, with a reservoir solution composed of 12–18% polyethylene glycol 8000, 100 mM sodium cacodylate, pH 7.0, and 25% glycerol. After the crystals grew over 3 weeks, they were removed by a loop and soaked in a solution composed of 12–18% polyethylene glycol 8000, 100 mM sodium cacodylate, pH 7.4, at room temperature. Abasic site cleavage reaction substrates were prepared by treating a 5'-32P-labeled uracil-containing 16-mer duplex DNA (5'-AGCGTCACTCC-3') with uracil DNA glycosylase (New England Biolabs) at 37 °C for 1 h. 100 nM pretreated duplex DNA was then incubated with 100 nM WT or N174C MutM in a reaction buffer of 200 mM NaCl, 50 mM Tris, pH 7.4, at room temperature. Aliquots of the reactions were removed periodically, quenched with an equal amount of 100 mM dithiothreitol in 95% formamide and 1 °C Tris/borate/EDTA buffer, and subjected to denaturing urea-PAGE electrophoresis and visualization on a phosphorimaging plate.

Fluorescence Polarization Experiments—A 3'-fluorescein-labeled DNA oligomer (5'-TGTTAGACCTGGACCG-fluorescein-3', purchased from Operon and gel-purified) was annealed 1:1 with either the tether-containing strand used in the XGC or the corresponding normal DNA without the N-ethylthio tether. The resulting normal DNA duplexes with or without tether were used in the WT and N174C MutM-DNA binding measurements in a buffer of 50 mM NaCl, 20 mM Tris, pH 7.4, and 0.1% β-mercaptoethanol. MutM at different concentrations (5 nM to 20 μM) was mixed with 10 nM DNA duplex in 96-well plates and equilibrated at room temperature. A SpectraMax M5 microplate reader (Molecular Devices) was used to monitor the changes in fluorescence polarization of the samples using excitation at 485 nm and emission at 530 nm. Dissociation constants were extracted by curve fitting of the changes in fluorescence polarization versus log of protein concentrations using Kaleidagraph 3.6. The experimental data points were measured in triplicate.

Molecular Dynamics Simulations of the DXL Effects—Systems with extrahelical oxoG or G bound to MutM were set up based on the x-ray crystal structure of LRCN174C or XGC, respectively. The positions of the hydrogen atoms were determined using the HBUILD facility in the CHARMM program version c33a2 (24, 25). For the DNA duplexes, we kept only the central 14 base pairs and discarded the flanking segments. The total charges of the resulting protein and DNA duplex systems were neutralized by placing Na+ ions (18 for un-cross-linked and 17 for cross-linked simulations) 4.5 Å away from the phosphorus atom along the line passing through the phosphorus atom and the midpoint of the nonbridging oxygens. All of the ordered waters observed in the x-ray structures were included, and the resulting systems were further solvated with a box of water molecules measuring 80.0 × 65.0 × 65.0 Å3. Each system was first energy-minimized to alleviate high energy contacts with a series of constraints and harmonic restraints and
then equilibrated with molecular dynamics (MD) stimulations at constant pressure (26) for 0.7 ns. The dimensions of each system measured 78.8 × 63.8 × 63.8 Å³ after the equilibration. Production MD simulations were performed for 5 ns for each system in which the last 1-ns simulations were used to determine the average structures. In the simulations, the leapfrog Verlet algorithm (27) was used with a 2-fs integration time step, and SHAKE (28) was applied to bonds involving hydrogens. The temperature was maintained at 298 K by coupling the system to the external thermal bath (29), and the volume was held constant. The all-atom CHARMM 27 force fields (30, 31) were used to represent the protein, ions, and nucleic acids, and the TIP3P model (32) was employed to represent water molecules. For the protein backbone dihedral angles, the CMAP correction (33) was applied. Periodic boundary conditions were used for the simulations. For the electrostatics, the particle mesh Ewald summation method (34) was used with a cut-off distance of 9 Å for the real space summation. The van der Waals interactions were evaluated with the same cut-off distance, and a shift function was used to smoothly turn off the interaction energy at the cut-off distance.

RESULTS

DXL Strategy for Stabilizing the Target Nucleoside in an Extruded Conformation—The presentation of a normal nucleobase to the MutM active site is likely to be a rare event having a fleeting lifetime. We sought to devise a strategy with which to capture this state and elucidate its structure. Previously, in the case of hOGG1, which is structurally unrelated to MutM, capturing an XGC was accomplished by a DXL strategy that interfered directly with base pairing of the target G to its complementary C (35). In the case of MutM, we have employed an alternative strategy involving conformational manipulation of the DNA backbone.

Computer simulations of the base extrusion processes indicated that there is a significant energetic barrier (~11 kcal/mol) for an intrahelical G to be extruded by MutM into a fully extrahelical and rotated conformation followed by insertion into the enzyme active site (14). Furthermore, the free energy of the fully extruded state for G bound to MutM is ~7 kcal/mol higher than for the initial intrahelical state. This, plus the fact that single-molecule sliding experiments estimated that the one-dimensional diffusion rate of MutM approaches that of the theoretical limit (36), essentially precludes the possibility that MutM extrudes and examines every normal nucleobase in its active site during the search for oxoG lesions. Additionally, MutM has no preference to bind any particular DNA sequence. These considerations suggested it would be necessary to provide an energetic driving force to stabilize G in the extrahelical conformation and to restrict the roaming range of the enzyme to the immediate vicinity of the target G. We reasoned that both of these objectives might be accomplished via DXL technology, which has previously been employed successfully to trap various other fleeting species along the MutM lesion search and extrusion pathway (14, 16).

Comparison of EC and IC structures (14) with those of the LRC (15) revealed, not surprisingly, that the most drastic structural differences brought about by extrusion of the target nucleoside are localized to the site of the target nucleoside, including the flanking 5’- and 3’- phosphates. We reasoned that we might be able to attach a phosphoramidate tether to one of these phosphates, thereby providing a point from which to tug on the DNA backbone at the site of the target nucleoside, thereby forcing it to adopt a conformation approximating that in the extrahelical state. Inspection of the previously reported structure of a MutM LRC (LRC1R2Y, where 1R2Y is the Protein Data Bank code) (15), in which oxoG is extrahelical and bound specifically in the enzyme active site having an ordered OCL, led to the identification of a key residue, Asn174, that employs its side chain amide to hydrogen bond with the 5’ phosphate of the target nucleoside (supplemental Fig. S1A). The distances separating Cy of Asn174 from the pro-S and pro-R nonbridging 5’-phosphate oxygens of oxoG in the LRC1R2Y are 4.3 and 3.9 Å, respectively. In an IC sequence-matched to LRC1R2Y (16), the corresponding distances are 7.1 and 5.3 Å for the respective pro-S and pro-R oxygens. From previous structures having an N-ethylthio-Cysj phosphoramidate DXL located at position 166 in MutM, we measured an average distance of 4.4 ± 0.3 Å between the cross-linker N and the Cys166S tether on the phosphate 5’- to oxoG (Figs. 2A and supplemental Fig. S2). Neither modification has a substantial effect (<2-fold) on the affinity of the MutM/DNA interaction (Fig. 2B). These results indicate that the component parts of the DXL, namely the N174C mutation in MutM and the N-ethylthio phosphoramidate in DNA, on the rate of the enzymatic repair reaction and on the strength of protein/DNA binding. The rate of enzyme-catalyzed oxoG and abasic site cleavage is barely affected by the presence of the N174C mutation and only modestly diminished by the presence of the N-ethylthio-bearing tether on the phosphate 5’- to oxoG (Figs. 2A and supplemental Fig. S2). Neither modification has a substantial effect (<2-fold) on the affinity of the MutM/DNA interaction (Fig. 2B). These results indicate that the component parts of the DXL, namely the N174C mutation in MutM and the N-ethylthio tether in DNA, do not significantly compromise the catalytic function of MutM.

Structure of a Control LRCN174C—Having the DXL Introduced—To validate the N174C cross-linking strategy and test whether the DXL perturbs the MutM-DNA structure in unintended ways, we determined the 1.7 Å structure of an LRCN174C having the DXL introduced (Table 1, and supplemental Figs. S1B and S3); this complex also bore the catalysis-inactivating E3Q mutation. In the structure of LRCN174C, electron density is plainly visible and interpretable for the atoms of the disulfide cross-link (Fig. 3C); although the DNA used in the DXL reaction consisted of a mixture of stereoisomers at the tethered 5’-phosphoramidate, the cross-linked product appears to bear only the S-configured phosphoramidate, thus indicating strong stereochemical selectivity in the cross-linking reaction. In other systems, including the case of Q166C MutM, we have invariably observed stereochemical mixtures in the crystallized cross-linked products.
The structure of the control DXLed LRC_{N174C} is nearly identical in structure to that of the un-DXLed LRC_{1R2Y} (root mean square deviation = 0.263 Å for all protein heavy atoms) and to other LRCs obtained using the Q166C DXL system (14, 16). Of particular importance, in the LRC_{N174C} structure, the target oxoG is fully extrahelical and correctly inserted into the MutM active site, with the OCL being well ordered (Fig. 3A). Despite the presence of a covalent cross-link to the 5′-phosphate, the DNA backbone conformation is nearly identical in LRC_{N174C} and LRC_{1R2Y} (Fig. 4A).

The most significant difference in the protein-DNA interface of the two LRC structures is the swiveling of the Arg264 side chain, which results in disengagement from its contacts to the oxoG 5′- and 3′-phosphates in LRC_{1R2Y} and establishment of new contacts to nearby DNA bases in LRC_{N174C} (supplemental Fig. S4). This change in the Arg264 contact repertoire is accompanied by a slight (1 Å) movement of the loop containing Arg264 (10-11 loop; supplemental Fig. S5). The reorganization of Arg264 appears to be driven by the presence of the DXL, specifically the need to avoid a steric clash between the sulfur atoms of the disulfide cross-link and the guanidinium head group of the amino acid.

Notwithstanding the minor adjustments at Arg264 and the β10-β11 loop, the overwhelming structural similarity between an un-DXLed LRC and one having a DXL introduced at Asn174 and the oxoG 5′-phosphate provides validation that this DXL system faithfully recapitulates physiologically relevant MutM/DNA interactions and therefore is suitable for the study of G presentation to the enzyme active site.

Structure of an Extrahelical G Complex, with a Normal G Being Presented to the MutM Active Site—Having validated the N174C DXL strategy by determining and analyzing the structure of LRC_{N174C}, we
moved on to using the same system to study G presentation to the MutM active site. We simply replaced the oxoG nucleobase in LRC_{N174C} with G, a difference of two atoms in the ∼40-kDa complex. We retained the catalysis-inactivating E3Q mutation, so as to avoid the possibility of MutM-catalyzed cleavage of the undamaged nucleobase over the extended time periods elapsed during crystallization and data collection.

The structure of the XGC, solved to 2.6 Å, revealed the target base present in both structures. Three MutM residues, Met77, Arg112, and Phe114, penetrate the DNA helix through the minor groove and stabilize the DNA bearing an extrahelical target base present in both structures. Three MutM residues, Met77, Arg112, and Phe114, penetrate the DNA helix through the minor groove and stabilize the DNA duplex (supplemental Fig. S6). Specifically, Phe^{114} inserts itself between the “estranged” C (opposite the extruded nucleobase), causing the target base pair to buckle and de-stack from the neighboring base pair. Arg^{112} hydrogen bonds with the extruded nucleobase and Phe^{114} penetrates the DNA helix through the minor groove and stabilizes the sharp bend in the DNA duplex (supplemental Fig. S6).

The XGC structure (Fig. 3B) shares many characteristic structural elements with the LRC_{N174C} structure (Fig. 3A). These features serve to stabilize the significantly deformed DNA bearing an extrahelical target base present in both structures. Three MutM residues, Met^{77}, Arg^{112}, and Phe^{114}, penetrate the DNA helix through the minor groove and stabilize the sharp bend in the DNA duplex (supplemental Fig. S6). Specifically, Phe^{114} inserts itself between the "estranged" C (opposite the extruded target base) and the 5'-neighboring nucleoside, causing the target base pair to buckle and de-stack from the neighboring base pair. Arg^{112} hydrogen bonds with the extruded C and, along with Met^{77}, fills in the void left behind by the extruded nucleobase. In addition, MutM contacts the backbone phosphates of both DNA strands in the region surrounding the target base to further enforce the bent DNA conformation (supplemental Fig. S4).

The direct hydrogen bonds involved in the LRC_{N174C} protein-DNA interface are preserved in the XGC; the side chains of Lys^{258}, Tyr^{242}, Gln^{3}, Lys^{60}, His^{74}, and Arg^{69} contact the DNA strand containing the target base, and the side chains of His^{93}, Lys^{113}, Asn^{92}, and Trp^{10} hydrogen bond to the complementary strand. A rigorous comparison of the water-mediated hydrogen bonds in the two structures is not possible because the XGC and LRC_{N174C} structures are not of the same resolution. However, in both cases, there seem to be relatively few ordered waters at the protein-DNA interface mediating indirect contacts, as compared with intrahelical
DNA-MutM structures (ICs and ECs) (14). The similarity between the hydrogen-bonding networks at the MutM-DNA interface in the LRCN\textsubscript{174C} and the XGC is expected, considering the similar protein and DNA backbone conformations in the two structures, most likely imposed by the engineered DXL (discussed above).

The main distinctions between the XGC and LRCN\textsubscript{174C} structures lie in the conformations of the target nucleobases and the positions of the MutM residues involved in stabilizing these conformations. The oxoG in the LRCN\textsubscript{174C} completely enters the active site and is stabilized by the OCL, forming multiple hydrogen bonds to the O-6 of the oxoG (amino Ns of Val\textsubscript{222}, Arg\textsubscript{223}, Thr\textsubscript{224}, and Tyr\textsubscript{225}, Fig. 5B). The oxoG is specifically contacted by the side chain of Gln\textsubscript{3} at the O-8 position and the carbonyl oxygen of Ser\textsubscript{220} at the protonated N-7. Although G bears the same O-6 functionality as oxoG, G differs from oxoG at C-8 and N-7, and this difference is sufficient to prevent the target G in the XGC from fully entering the active site and engaging the OCL; instead, the target G lies parallel to the minor groove floor of the DNA (Fig. 3B). The orientation of the extruded G is almost perpendicular to that of the extrahelical oxoG, and the OCL in the XGC is disordered throughout most of its length (Fig. 3B and supplemental Fig. S3). Interestingly, in the complexes of MutM interacting with an intrahelical oxoG or normal base (ECs or ICs), the entire OCL (residues 216–237) is disordered, but in the XGC, 12 more residues of the OCL are ordered than in the intrahelical complexes (Fig. 3D). The disordered region of the OCL (residues 222–231) comprises the tip of the loop, which contains residues that interact directly with extrahelical oxoG. In contrast to the extensive contacts between the target oxoG and the MutM active site in the LRCN\textsubscript{174C}, only a few hydrogen bonds appear to be established between MutM and the extrahelical G in the XGC: the side chain of Glu\textsubscript{78} and the carbonyl oxygen of Met\textsubscript{77} with N-2 and Tyr\textsubscript{242} with N-7, none of which is specific for G versus oxoG (Fig. 5A).

Overlaying the 5'-phosphate groups from the extrahelical G and the oxoG in the active site (supplemental Fig. S7) reveals that the center of mass for the deoxyribose relative to the backbone phosphate is quite similar for the two target bases, but the glycosidic bond angle between the base and the sugar (the O-4'–C-1'–N-9–N-3 dihedral angle) differs significantly (162° for the G and 96° for the oxoG). Although the active site oxoG is in the syn conformation (Fig. 6B), the extrahelical G remains close to the anti conformation (Fig. 6A). Conversion of the extrahelical G conformation to the active site oxoG conformation requires (i) pseudo-rotation of the deoxyribose (i.e. downward movement of the C-3' with respect to the center of the sugar) to relax the sugar to the C-2'-endo conformation, and (ii) rotation of the glycosidic bond to assume the syn conformation, which orients O-6 of the base in the proper direction for interactions with the OCL (supplemental Fig. S7). In summary, the XGC structure represents MutM examining a fully extrahelical but not fully rotated G in an “exo-site” outside a partially assembled active site.

Simulating the Structure of an un-DXLed G Complex—Although the foregoing analysis established that the presence of the N174C DXL in the LRCN\textsubscript{174C} complex does not perturb the DNA structure at the interface with oxoG, it does have a slight effect on the protein conformation, specifically of Arg\textsubscript{264} and the β10-β11 loop. An un-DXLed XGC is obviously inaccessible experimentally, but we reasoned that because the effects of DXL are modest and localized, we could be confident in the use
of equilibrium MD simulations to remove the effect of DXL and provide a close semblance of a physiologic intermediate with G being presented to the MutM active site. A series of MD simulations was performed starting with the crystal structure of the DXLed XGC, and the results were compared with a second series of MD simulations carried out in the absence of the DXL. Corresponding MD simulations were performed for the LRCN174C structure in the presence and absence of the DXL, again as a control. An averaged structure for each system was determined from each run of MD simulations. As expected, the MD-averaged structures for the DXLed XGC and LRC systems remained close to those of the corresponding crystal structures (heavy atom root mean square deviation = 1.292 Å for the XGC and 0.830 Å for the LRC). The global structures generated via the un-DXLed MD simulations also remained close to those of the relevant crystal structures (heavy atom root mean square deviation = 0.800 Å for the XGC and 0.648 Å for the LRC), consistent with the notion that the effect of cross-linking to N174C is restricted to minor structural adjustments.

Close inspection of the MD-averaged XGC structures, in comparison with the crystal structure, reveals that the extruded G of the un-DXLed system (Fig. 7D) is oriented slightly differently from that of the simulated DXLed system (Fig. 7C) and the crystal XGC structure (Fig. 7A). One notable difference lies in the glycosidic bond angle of the extruded G, which is 162° in the crystal XGC structure, 169° for the DXLed MD-averaged structure, and 244° for the un-DXLed MD-averaged structure. As a reference, for the LRC systems with almost identical oxoG configurations for DXLed and non-DXLed crystal structures, the glycosidic bond angle of the oxoG in the active site is 72° for the DXLed MD simulations and 62° for the un-DXLed simulations, respectively, whereas that for the crystal structure is 99° (supplemental Fig. S8). It is particularly noteworthy that previous potential of mean force simulations (14), carried out by us before the present XGC was available, have suggested the existence of an extrahelical intermediate state in G extrusion by MutM (Fig. 7B); the orientation of the extruded G in the potential of mean force simulations is almost identical to that of the G from the present un-DXLed MD simulations, i.e. the glycosidic bond angle for the potential of mean force extrahelical intermediate structure is ~240°, essentially the same as the angle from the present un-DXLed simulations.

A second point of difference between the DXLed and un-DXLed structures lies in the conformation of Arg264 and the β10-β11 loop. As mentioned above, the positions of these elements shift slightly upon DXL introduction, as evidenced by detailed comparison of the crystal structures of (un-DXLed) LRC1R2Y and (DXLed) LRCN174C. Computational removal of the DXL in LRCN174C restores Arg264 and the β10-β11 loop to nearly identical conformations as seen in LRC1R2Y (supplemental Fig. S8). This restoration can be understood as resulting from removal of the unfavorable steric interaction between the DXL moiety and the side chain of Arg264, a residue that is part of the β10-β11 loop. The forces that cause the glycosidic bond adjustment are less obvious from the available data but may also reflect the influence of Arg264, specifically the presence or absence of its interactions with the 5’ and 3’ phosphates of the target G.

Taken together, the MD simulation results support the notion that the presence of the DXL in the XGC does not significantly alter the protein/DNA interface in the XGC, and what minor differences do exist are readily removed through the MD simulations so as to provide an otherwise inaccessible physiologic extrusion intermediate.

**DISCUSSION**

*Trapping Transient Intermediates in Protein/DNA Interaction Systems Using DXL*—Intermolecular disulfide cross-linking of protein-DNA complexes, first employed to enable structure determination of the HIV reverse transcriptase-prim-er-template complex (37), has since been used in a variety of systems to trap fleeting or rarely formed intermediates and to restrict the roaming range of proteins on DNA, thereby increas-
Structure of MutM with an Extrahelical Guanine

...ing the homogeneity of otherwise inhomogeneous complexes (14, 16, 35, 38–41). The process of forming an intermolecular disulfide bond between DNA and protein is typically carried out under conditions in which thiol-disulfide interchange is kinetically reversible and can thereby attain an equilibrium position (supplemental Fig. S1B). Although some thermodynamic driving force in the cross-linking reaction is provided by the high local concentration of the engineered Cys on the protein and thiol tether on the DNA with respect to each other, this effect is modest and therefore insufficient under equilibrating conditions to drive the formation of highly strained species. In this way, DXL differs from kinetically irreversible covalent conjugation methods such as photocross-linking, which involve highly reactive intermediates that can drive the formation of strained species. By virtue of formative equilibration, DXL is predisposed toward trapping structures that are readily accessible energetically and are therefore likely to be substantially represented under physiologic conditions.

In the present case, a pair of interacting thiols was engineered into the interface between MutM and an undamaged duplex DNA, with a sufficiently large gap between the thiol partners that either the protein, DNA, or both would have to undergo structural remodeling to form a disulfide linkage. We found that indeed the protein and DNA readily reacted to form a DXL. Inspection of the structure of this complex revealed that the structural remodeling entailed extrusion of the normal G nucleoside 3’- to the N-ethylthio-tethered phosphoramidate (i.e. the G having the tether on its 5’-phosphate); very little structural remodeling appears to have taken place on the part of the protein.

Validation of the structure of this MutM XGC was provided by generating a control lesion recognition complex (LRCN174C) that differed from the XGC only in having a target oxoG in place of G, a difference of two atoms. The DNA conformation at the site of the extruded target base in the DXLed LRCN174C control structure proved to be nearly identical to that of un-DXLed LRC structures, and only minor perturbations attributable to the cross-link were evident. These perturbations in LRCN174C were found to be readily rectified via MD simulations. Having thus also validated the MD simulations using the LRC structures, we applied the computational rectification procedure to the XGC structure, thereby producing a hybrid crystallographic/computational structure that provides a view of MutM extruding an undamaged G nucleoside from DNA and presenting it for insertion into the extrahelical enzyme active site.

Discrimination against Normal Extrahelical DNA Bases by MutM—Because the structures of LRCN174C and XGC and of the MD-rectified versions thereof differ only in whether they possess oxoG or G, respectively, at the target nucleoside and because the differences are not ascribable to crystal packing factors, we are confident in the conclusion that the observed structural differences result from the two-atom structural divergence in the target nucleoside. Whereas oxoG undergoes full insertion into the MutM active site, causing the OCL to become ordered and engage the Hoosteen face of the nucleoside, G undergoes only partial insertion, and the OCL fails to become fully ordered.

What are the structural features of G relative to oxoG that cause the undamaged nucleoside to be rejected by the MutM active site? To gain insight into this question, we modeled the structure of a MutM-DNA complex having an extruded G fully inserted into the MutM active site (Fig. 5C); this was done in a straightforward manner by changing oxoG to G in LRCN174C. As expected, the multiple contacts between the OCL and the O-6 of oxoG would appear to be preserved with O-6 of G (Fig. 5, compare C and D), and thus O-6 contacts are unlikely to play a role in discriminating G from oxoG. On the other hand, oxoG and G do differ in their protonation state at N-7, and this is likely to affect the interaction with the main chain carbonyl of Ser220. Specifically, the attractive hydrogen bonding interaction between the Ser220 amide carbonyl and the oxoG N-7-hydrogen would be replaced by a repulsive interaction when presented with the lone pair of electrons on N-7 of G (Fig. 5, compare B with C). Indeed, studies with analogs of oxoG have shown that a protonated N-7 position is essential for lesion recognition by MutM (and hOGG1) (42). Although the OCL could in principle adjust to avoid the repulsive interaction with N-7 of G, such an adjustment would possibly dislodge the contacts to O-6.

In addition, in the unpaired state, oxoG has a strong preference for the syn glycosidic bond conformation, whereas an unpaired G does not prefer syn over anti (43). Full engagement of oxoG in the MutM active site requires the syn conformation, and this factor therefore provides an additional thermodynamic bias toward oxoG and against G recognition by the MutM active site. In summary, a combination of differences in N-7 protonation state and glycosidic torsion angle preferences appear to provide the structural cues that the MutM active site uses to distinguish an oxoG lesion from its closely related normal counterpart G.

In work published elsewhere (14), we have found that mutations in the OCL that locally alter its conformation near the residues involved in oxoG recognition completely abrogate base excision repair of oxoG while leaving intact the ability of the enzyme to catalyze β-lyase degradation of abasic sites. Thus, proper ordering and positioning of the OCL over the substrate appears to be essential for catalysis of base excision by MutM. Presentation of a G residue to the active site also prevents the OCL from attaining its proper position; this, combined with the higher intrinsic glycosidic bond stability of G versus oxoG (42), can be safely assumed to make even a forcibly extrahelical G a poor substrate for MutM.

This work implies that MutM, like hOGG1, has a late checkpoint in its nucleoside extrusion pathway that consists of acceptance or rejection of an extruded substrate nucleoside on the basis of active site complementarity. The existence of this checkpoint prevents MutM from accidentally cleaving G, which is present in vast excess over oxoG. How frequently is this catalytic checkpoint called upon to function? One potential source of extrahelical G residues for MutM to bind is those that have been produced by random thermal fluctuations in the DNA. Normal base pairs in DNA have a lifetime of 5–100 ms as...
measured by NMR imino proton exchange experiments (44), but this figure reflects the rate of breathing and not that of extrusion. Breathing involves minimal loss of base stacking interactions and little increase in solvent exposure of hydrophobic surface area, whereas spontaneous extrusion entails both of these energetically costly occurrences; therefore, the rate of spontaneous nucleobase extrusion from DNA is orders of magnitude slower than the breathing rate (45). The near barrier-free one-dimensional diffusion of MutM along DNA \((3.5 \pm 0.6 \times 10^{3} \text{ base pair}^2/\text{s})\) implies that chance encounters of MutM with spontaneously extrahelical nucleosides are extremely rare (36), to the point of being negligible. Another source is G residues whose extrusion from DNA has been accelerated through interaction with MutM. Studies in our laboratories have demonstrated that when MutM binds even undamaged DNA, it induces significant helical bending and de-stacking at, and buckling of, target base pairs (16). This is associated with a lowering of the energy barrier to extrusion. The barrier for o xoG extrusion is lowered by 14 kcal/mol more than that for G, but even for G, MutM-assisted extrusion is much faster than in naked DNA (14). A calculated energy barrier to MutM-assisted G extrusion of 11 kcal/mol suggests that the enzyme extrudes G at a rate of \(\sim 1 \times 10^{-8} \text{ ps}^{-1}\). This is frequent enough to be deleterious to cells; hence we conclude that the catalytic checkpoint serves a vital biological function.

Finally, structures of DNA glycosylases having partially extruded undamaged bases have also been observed previously with hOGG1 (35, 46) and uracil DNA glycosylase (47). As observed for the case of MutM XGC, a common feature of these structures is the few contacts between the target nucleobase and the enzyme exo-site. Therefore, stalling partially extruded normal nucleobases in exo-sites might be a general mechanism of lesion recognition at the late stage of base extrusion for DNA glycosylases.

**Implications for the MutM Base Extrusion Pathway**—The present structure of MutM attempting to present an undamaged nucleobase to the active site provides a representation of a formerly cryptic state along the base extrusion pathway of MutM; as such, it complements the previously reported structures of MutM-DNA complexes at other stages of lesion recognition and base extrusion (IC, EC, and LRC; Fig. 1B) (14–16). As shown in the previous computational studies, the base extrusion pathway of MutM involves three steps: (i) extrusion of the nucleobase from the intrahelical base pair to a fully extrahelical position through the minor groove, (ii) rotation of the nucleobase to a syn conformation, and (iii) further extrusion of the base to settle into the enzyme active site. Compared with the fully extruded o xoG in the LRC, the G in the XGC has only completed the first step of the extrusion process but has not completed the second or third step. Using MD simulations, the XGC can be readily relaxed into a structure very close to that of a previously calculated local free energy minimum state \((Z = 1.0)\) (14) on the two-dimensional free energy surface for the extrusion of G assisted by MutM. Therefore, consistent with the computational model of base extrusion by MutM, the XGC structure suggests that complete rotation of the nucleobase is a late event following the extrusion of the base to a fully extrahelical position. Furthermore, ordering of the OCL accompanies the extrusion of the base, and the OCL does not assume a single conformation until full extrusion of the target base occurs.

**Model for Overall Damage Recognition by MutM**—The complete set of MutM-DNA structures along with the currently available biochemical and biophysical data suggest the following model for lesion recognition and discrimination against normal nucleobases by MutM: MutM scans along the DNA rapidly in a highly redundant mode by one-dimensional Brownian motion. During this process, the protein interrogates the duplex, bending it and buckling target base pairs more or less indiscriminately. Upon encountering an intrahelical o xoG, the free energy barrier to extrusion of the target lesion is selectively lowered as the result of steric interactions between o xoG and the bent DNA backbone induced by interactions with MutM, and the lesion therefore is extruded in preference to undamaged bases. Once extruded, o xoG rotates into the syn conformation and is captured and stabilized by the OCL in the active site. Whenever MutM accidentally promotes the extrusion of G, the undamaged nucleoside reaches the exo-site but fails to enter the active site fully and cause the OCL to lock down over it. Absent proper assembly of a functional catalytic apparatus, the partially extruded G fails to undergo base excision and quickly collapses back into the intrahelical state. Because the other three undamaged nucleobases have very different chemical compositions from o xoG, discrimination against these nucleobases by MutM is not expected to be as challenging as with G. In conclusion, through a combination of intrahelical and extrahelical lesion recognition mechanisms, MutM efficiently locates the rare o xoG sites and avoids being overwhelmed by the more abundant normal nucleobases in the genome.

**Acknowledgments**—We are grateful to Annie Héroux of the Macromolecular Protein Crystalllography mail-in program from National Synchrotron Light Source for collection of the LRCN174C data set and to the staff from the 19-ID Beamline of the Argonne Photon Source for assistance in data collection. We also thank members of the Karplus and Verdine Groups for helpful advice.

**REFERENCES**

1. Lindahl, T. (1993) *Nature* 362, 709–715
2. Barnes, D. E., and Lindahl, T. (2004) *Annu. Rev. Genet* 38, 445–476
3. Zharkov, D. O. (2008) *Cell Mol. Life Sci.* 65, 1544–1565
4. Hitomi, K., Iwai, S., and Tainer, J. A. (2007) *DNA Repair* 6, 410–428
5. Tchou, J., Bodepudi, V., Shibutani, S., Antoshechkin, I., Miller, J., Grollman, A. P., and Johnson, F. (1994) *J. Biol. Chem.* 269, 15318–15324
6. Grollman, A. P., and Moriya, M. (1993) *Trends Genet.* 9, 246–249
7. Shibutani, S., Takeshita, M., and Grollman, A. P. (1991) *Nature* 349, 431–434
8. Weiss, J. M., Goode, E. L., Ladiges, W. C., and Ulrich, C. M. (2005) *Mol. Carcinog.* 42, 127–141
9. Cadet, J., Douki, T., Gasparutto, D., and Ravanat, J. L. (2003) *Mutat. Res.* 531, 5–23
10. Plum, G. E., Grollman, A. P., Johnson, F., and Breslauer, K. J. (1995) *Biochemistry* 34, 16148–16160
11. Lipscomb, L. A., Peek, M. E., Morningstar, M. L., Verghis, S. M., Miller, E. M., Rich, A., Essigmann, J. M., and Williams, L. D. (1995) *Proc. Natl. Acad. Sci. U.S.A.* 92, 719–723
12. Bowman, B. R., Lee, S., Wang, S., and Verdine, G. L. (2008) *Structure* 16, 1166–1174
Structure of MutM with an Extrahelical Guanine

13. Kunkel, T. A., and Erie, D. E. (2005) Annu. Rev. Biochem. 74, 681–710
14. Qi, Y., Spong, M. C., Nam, K., Banerjee, A., Jiralerspong, S., Karplus, M., and Verdie, G. L. (2009) Nature 462, 762–766
15. Fromme, J. C., and Verdie, G. L. (2003) J. Biol. Chem. 278, 51543–51548
16. Banerjee, A., Santos, W. L., and Verdie, G. L. (2006) Science 311, 1153–1157
17. Fromme, J. C., and Verdine, G. L. (2002) Nat. Struct. Biol. 9, 544–552
18. Otwonowski, Z., and Minor, W. (1997) Methods Enzymol. 276, 307–326
19. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Acta Crystallogr. D 54, 905–921
20. Emsley, P., and Cowtan, K. (2004) Acta Crystallogr. D 60, 2126–2132
21. Brünger, A. T. (1997) Methods Enzymol. 277, 366–396
22. Schüttelkopf, A. W., and van Aalten, D. M. (2004) Acta Crystallogr. D 60, 1355–1363
23. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sauter, N. K., and Terwilliger, T. C. (2002) Acta Crystallogr. D 58, 1948–1954
24. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) J. Comput. Chem. 4, 187–217
25. Brooks, B. R., Brooks, C. L., 3rd, Mackrell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresh, S., Caflisch, A., Caves, L., Cui, Q., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) J. Comput. Chem. 30, 1545–1614
26. Andersen, H. C. (1980) J. Chem. Phys. 72, 2384–2393
27. Allen, M. P., and Tildesley, D. J. (1989) Computer Simulation of liquids, pp. 71–109, Oxford University Press, USA
28. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) J. Comput. Phys. 23, 327–341
29. Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) J. Chem. Phys. 81, 3682–3690
30. Mackrell, A. D., Jr., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchmir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reinh, W. E., Ill, Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) J. Phys. Chem. B 102, 3586–3616
31. Fopolpe, N., and Mackrell, A. D. (2000) J. Comput. Chem. 21, 86–104
32. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) J. Chem. Phys. 79, 926–935
33. MacKerell, A. D., Jr., Feig, M., and Brooks, C. L., 3rd (2004) J. Am. Chem. Soc. 126, 698–699
34. Darden, T., York, D., and Pedersen, L. (1993) J. Chem. Phys. 98, 10089–10092
35. Banerjee, A., Yang, W., Karplus, M., and Verdie, G. L. (2005) Nature 434, 612–618
36. Blaney, P. C., van Oijen, A. M., Banerjee, A., Verdie, G. L., and Xie, X. S. (2006) Proc. Natl. Acad. Sci. U.S.A. 103, 5752–5757
37. Huang, H., Chopra, R., Verdie, G. L., and Harrison, S. C. (1998) Science 282, 1669–1675
38. Verdie, G. L., and Norman, D. P. (2003) Annu. Rev. Biochem. 72, 337–366
39. Corn, J. E., and Berger, J. M. (2007) Structure 15, 773–780
40. Duguid, E. M., Mishina, Y., and He, C. (2003) Chem. Biol. 10, 827–835
41. Yang, C. G., Yi, C., Duguid, E. M., Sullivan, C. T., Jian, X., Rice, P. A., and He, C. (2008) Nature 452, 961–965
42. Hamm, M. L., Gill, T. J., Nicolson, S. C., and Summers, M. R. (2007) J. Am. Chem. Soc. 129, 7724–7725
43. Jiang, Y. H., Godward, W. A., 3rd, Noyes, K. T., Sowers, L. C., Hwang, S., and Chung, D. S. (2002) Chem. Res. Toxicol. 15, 1023–1035
44. Dornberger, U., Leijon, M., and Fritzscbe, H. (1999) J. Biol. Chem. 274, 6957–6962
45. Priyakumar, U. D., and Mackrell, A. D., Jr. (2006) J. Am. Chem. Soc. 128, 678–679
46. Banerjee, A., and Verdie, G. L. (2006) Proc. Natl. Acad. Sci. U.S.A. 103, 15020–15025
47. Parker, J. B., Bianchet, M. A., Krosky, D. J., Friedman, J. I., Amzel, L. M., and Verdine, G. L., 3rd (1982) J. Am. Chem. Soc. 104, 35–44
48. Banerjee, A., Yang, W., Karplus, M., and Verdie, G. L. (2005) Nature 434, 612–618
49. Fromme, J. C., and Verdine, G. L. (2002) Nat. Struct. Biol. 9, 544–552