Chemistry & Biology, Vol. 10, 827–835, September, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.chembiol.2003.08.007

How Do DNA Repair Proteins Locate Potential Base Lesions? A Chemical Crosslinking Method to Investigate O^6-Alkylguanine-DNA Alkyltransferases

Erica M. Duguid, Yukiko Mishina, and Chuan He*
Department of Chemistry
University of Chicago
5735 South Ellis Avenue
Chicago, Illinois 60637

Summary

O^6-alkylguanine-DNA alkyltransferases directly reverse the alkylation on the O^6 position of guanine in DNA. This group of proteins has been proposed to repair the damaged base in an extrahelical manner; however, the detailed mechanism is not understood. Here we applied a chemical disulfide crosslinking method to probe the damage-searching mechanism of two O^6-alkylguanine-DNA alkyltransferases, the *Escherichia coli* C-Ada and the human AGT. Crosslinking reactions with different efficiency occur between the reactive Cys residues of both proteins and a modified cytosine bearing a thiol tether in various DNA probes. Our results indicate that it is not necessary for these proteins to actively flip out every base to find damage. Instead they can locate potential lesions by simply capturing a lesioned base that is transiently extrahelical or sensing the unstable nature of a damaged base pair.

Introduction

Cellular DNA is constantly subjected to modifications by intracellular and extracellular chemicals, which can result in covalent changes [1, 2]. The heterocyclic bases of DNA are principle targets that can be modified by a variety of chemicals such as reactive oxygen species, alkylating reagents, and even water [2]. Much of the damage, if not repaired, will alter genetic information and cause mutagenic consequences. Nearly all organisms evolve proteins to repair these lesions. For example, several base excision repair DNA glycosylase enzyme families are used to remove mutagenic lesions, such as 8-oxoguanine and 3-methyladenine, in the first step of the repair [1, 2]. In these cases, the damaged bases are flipped out of the DNA double helix and cleaved in the active site pockets of these enzymes [3–5]. Other repairing modes have also been identified; for instance, it was found that the methyl groups on 1-methyladenine and 3-methylcytosine lesions are directly eliminated by the AlkB protein through a novel oxidative demethylation mechanism [6–9].

O^6-alkylguanine-DNA alkyltransferases are another family of proteins that directly remove alkylation adducts on DNA bases [10–14]. They repair the alkylation damage occurring on the O^6 position of guanine or the O^4 position of thymine via irreversible transfer of the alkyl group to a nucleophilic Cys residue (Figure 1). This activity is an important component for cellular resistance to the toxic and mutagenic effects of alkylation damage since the deleterious modification that is corrected by this family of proteins, O^6-alkylguanine, represents one of the most mutagenic lesions known. This unrepaired damage can lead to the transition mutation of G•C to A•T. Because of the potential hazard of this mutation, O^6-alkylguanine-DNA alkyltransferase is widely conserved; and the best-characterized proteins in this family are the C-terminal domain of Ada protein (C-Ada) from *Escherichia coli* and the human O^6-alkylguanine-DNA alkyltransferase (hAGT). Human AGT is homologous to C-Ada and can remove a variety of alkyl substitutions, including very large organic groups on the O^6 position of guanine and O^4 position of thymine [15, 16]. Treatment of cancer patients with alkylating antitumor agents induces the expression of hAGT by an unknown mechanism. The presence of the increased amounts of hAGT has been identified as the leading cause for the resistance of tumor cells to certain chemotherapeutic alkylating agents [17–19].

X-ray crystal structures of the 178 amino acid *E. coli* C-Ada and truncated human AGT have been obtained [20–22]. The structures show that the reactive Cys residue is hosted in a buried active site in both proteins. An "arginine finger" close to a DNA binding helix-turn-helix motif was found which could be used to extrude damaged nucleotides from duplex DNA. This observation together with subsequent mutation studies suggest that the damaged base is repaired extrahelically by hAGT. Despite these previous studies, the detailed molecular basis for damage recognition and damage repair for the O^6-alkylguanine-DNA alkyltransferase family remains unclear, primarily due to the lack of a protein/DNA complex structure.

How proteins that repair damaged bases extrahelically search for potential base lesions is a long-standing question [23]. Available structures of protein/DNA complexes [3–5, 24–29] have revealed that the protein gains access to its substrate by flipping out and inserting the base into an active site pocket. These observations led to speculation that the proteins use a similar mechanism to search for damage [23]. Namely, a base-specific DNA repair protein would flip out every base and detect potential lesions by checking it in its substrate binding pocket, as shown in Figure 2A. Thus, the protein migrates along DNA, checking every base in its active site until it locates the damaged base. The protein could also operate by two other mechanisms that do not involve flipping out every base. One of these mechanisms suggest that the protein could scan through the duplex DNA and directly detect an adduct or sense the distortion caused by a damaged base that is intrahelical, as illustrated in Figure 2B. The third mechanism does not require the protein to actively search for lesions. Base modifications incurred from damage could alter the structure of the base-pairing interface, causing the damaged base incapable of forming a stable Watson-Crick base pair with the opposite base on the complementary

*Correspondence: chuanhe@uchicago.edu

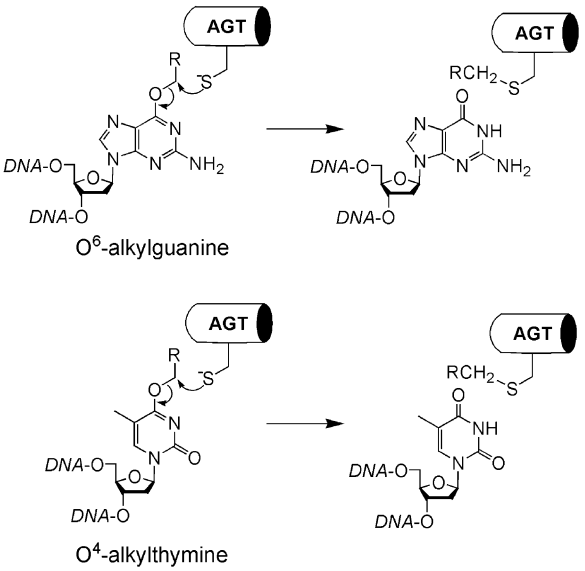

Figure 1. DNA Repair Functions Performed by O^6-Alkylguanine-DNA Alkyltransferases

strand. The damaged base could then rotate out of the duplex DNA structure and become extrahelical. Base repair proteins can capture the damaged base that is "flipped out" and perform the repair function, as indicated in Figure 2C.

These three mechanisms have never been directly confirmed since there are no reliable methods to study this problem. Understanding how nature has evolved strategies to locate a single modification on a base among billions of atoms in a genome is of fundamental interest. We report here a chemical crosslinking approach to help elucidate the mechanisms employed by members of the O^6-alkylguanine DNA alkyltransferase family, *E. coli* C-Ada and human AGT, in their search for potential lesions in the genome. Our results may have general implications for other base repair and modification systems. The method also allows us to trap covalently linked protein/DNA complexes for future structural studies.

Results and Discussion

A Disulfide Crosslink Strategy

E. coli C-Ada and human AGT use their corresponding Cys residues (Cys139 in truncated C-Ada and Cys145 in hAGT) to remove the alkyl groups on either the O^6 position of guanine or the O^4 position of thymine (Figures 1 and 3A). Both purine and pyrimidine substrates can be recognized and processed by these proteins [11, 16, 30]. The proteins are apparently quite flexible since they can remove alkyl groups that are at various distances (the O^6 position of guanine versus the O^4 position of thymine) from the reactive Cys residues. This flexibility may allow other modified bases to access the same substrate binding pockets and react with the Cys residues. Therefore, a reactive group on the 4 position of a modified pyrimidine could be positioned in close proximity to Cys139 in C-Ada (or Cys145 in hAGT), and the reactive moiety could potentially react with the Cys residue. To test this hypothesis, we employed a disulfide crosslinking strategy, starting with C-Ada, which has been used previously to effectively trap labile protein/DNA complexes under equilibrium conditions [31–33]. Oligonucleotides having a specifically modified cytosine base (C*) with a thiol tether introduced at the N^4 position were prepared, as shown in Figures 3B and 3C. It was hoped that the modified cytosine could be inserted into the substrate binding pocket of C-Ada; and then Cys139 could attack the tethered disulfide group to form a disulfide crosslink between the protein and the modified DNA (Figure 3B). We chose to use the thiol-tethered cytosine base because it can be readily and specifically introduced into DNA [34–36]. Adding a tether to the O^6 position of a guanine or the O^4 position of a thymine would be unsuitable since an alkyltransferase would simply remove the tether instead of covalently linking to it.

Double-stranded oligonucleotides with a modified cytosine base C* opposite G, A, and T (DNA-2G, -2A, and -2T in Figure 3C) were prepared to probe the damage-searching mechanism of C-Ada. The C* in DNA-2G is held relatively tightly in the duplex DNA structure by Watson-Crick hydrogen bonding to the opposite base G. With bases A or T opposite C* in DNA-2A or DNA-2T, the C* could sample the extrahelical conformation more often due to its significantly weakened ability to

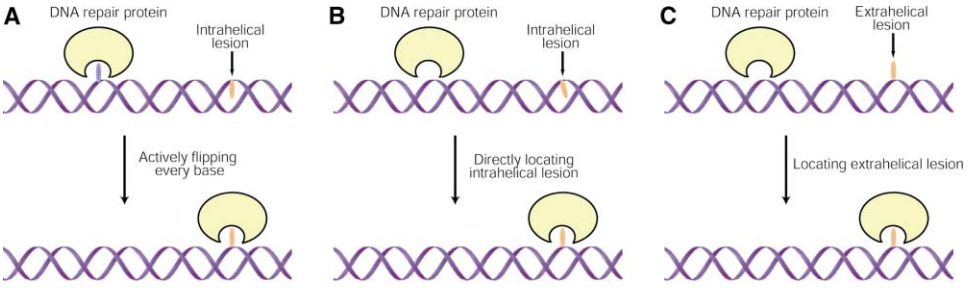

Figure 2. Proposed Damage-Searching Mechanisms for DNA Repair Proteins that Process Damaged Bases Extrahelically

(A) An active damage-searching mechanism. In this mechanism, the protein flips every base out and checks it in its active site pocket until the lesion is located.

(B) Proposed mechanism of damage searching by a DNA repair protein by directly detecting adducts that are hidden intrahelically.

(C) Proposed mechanism of damage searching by a DNA repair protein by simply capturing a transiently extrahelical lesion.

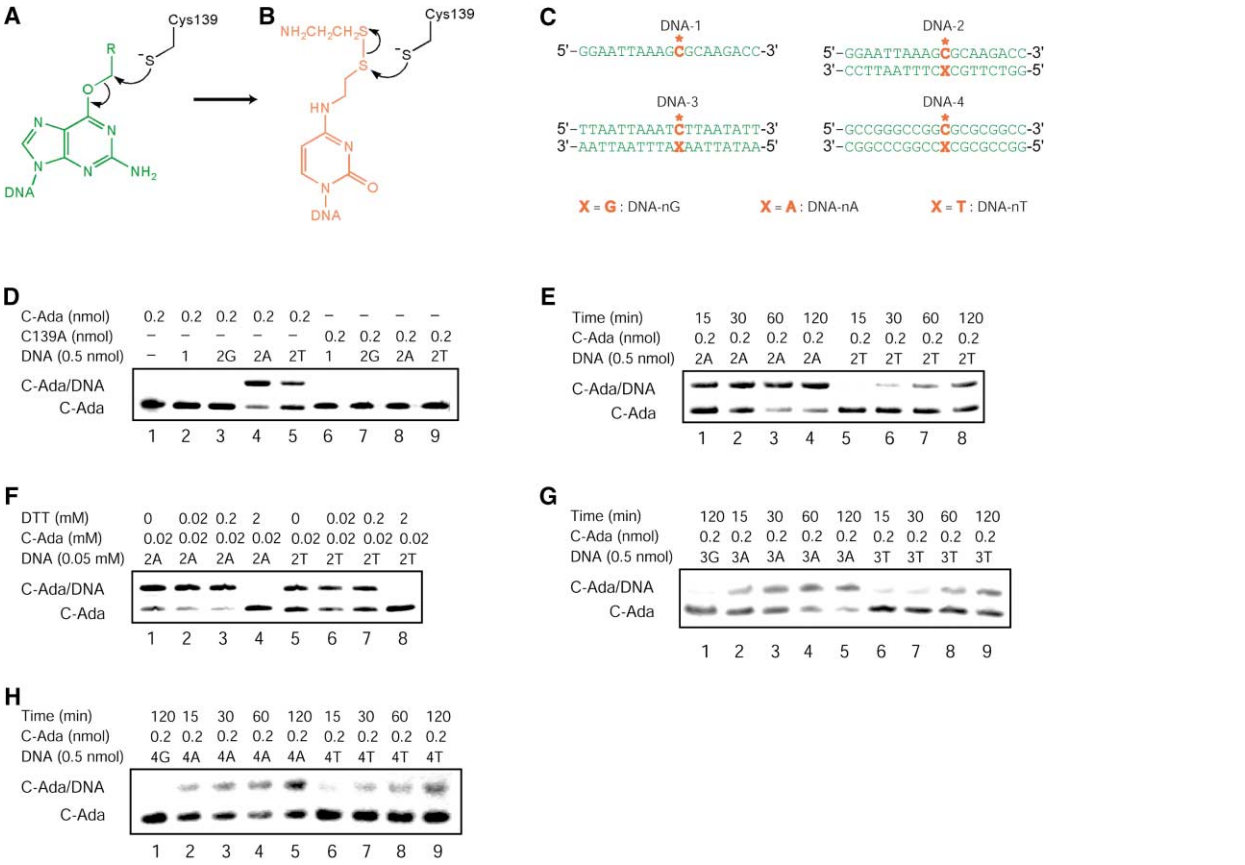

Figure 3. Disulfide Crosslinking between C-Ada and DNA-1, -2G, -2A, and -2T

(A) Residue Cys139 of C-Ada removes an alkyl group on the O^6 position of guanine in a duplex DNA.

(B) Replacement of O^6-alkylguanaine with a modified cytosine (C*) bearing an O^4-thiol tether in a duplex DNA provides a reactive disulfide group that can be attacked by Cys139 of C-Ada.

(C) Structures of oligonucleotides used in crosslinking. A single-stranded DNA (DNA-1) and three different sequences of double-stranded DNA (DNA-2, DNA-2A, and DNA-2T, with normal, AT-rich, and GC-rich composition) were used for crosslinking reactions. There are three variations of each double-stranded sequence having G, A, or T opposite C* denoted as DNA-nG, DNA-nA, and DNA-nT, respectively. For example, DNA-2G has a C*•G base pair in DNA-2 sequence, DNA-2A has a C*•A base pair in DNA-2 sequence, and DNA-2T has a C*•T base pair in DNA-2 sequence.

(D) SDS gel analysis of the crosslinking reaction between C-Ada and DNA-1, -2G, -2A, and -2T. All reactions were performed by incubating 20 μM of C-Ada and 50 μM of DNA at 4°C. Lane 1 is a size standard for C-Ada. Lanes 2–5, crosslinking results between C-Ada and DNA-1, -2G, -2A, and -2T after 24 hr incubation. Disulfide crosslinking between the protein and DNA results in the appearance of a new band having retarded mobility. Lanes 6–9, controls with C-Ada C139A mutant (note the lack of crosslinking).

(E) Time course of crosslinking reactions. Lanes 1–4, time course of crosslinking reaction between C-Ada and DNA-2A. Lanes 5–8, time course of crosslinking reaction between C-Ada and DNA-2T.

(F) Effect of external thiol (DTT) on the reactions. Lanes 1–4, crosslinking results between C-Ada and DNA-2A in the presence of varying amounts of DTT after 2 hr incubation. Lanes 5–8, crosslinking results between C-Ada and DNA-2T in the presence of varying amounts of DTT after 2 hr incubation.

(G) Time course of crosslinking reactions with AT-rich sequence. Lane 1, no product was observed for the crosslinking reaction of C-Ada and DNA-3G after 2 hr incubation. Lanes 2–5, time course of crosslinking reaction between C-Ada and DNA-3A. Lanes 6–9, time course of crosslinking reaction between C-Ada and DNA-3T.

(H) Time course of crosslinking reactions with GC-rich sequence. Lane 1, no product was observed for the crosslinking reaction of C-Ada and DNA-2TG after 2 hr incubation. Lanes 2–5, time course of crosslinking reaction between C-Ada and DNA-4A. Lanes 6–9, time course of crosslinking reaction between C-Ada and DNA-4T.

base pair with the mismatched base on the complementary strand [37, 38]. If C-Ada flips out every base during damage searching, as shown in Figure 2A, it should crosslink with all three DNA. If it directly detects an adduct hidden on an intrahelical base in the duplex DNA structure and then flips the base out for the repair (Figure 2B), the protein should crosslink with DNA-2G. It may also crosslink with DNA-2A and DNA-2T. If C-Ada does not flip out every base but instead captures the damaged base that is already extrahelical as indicated in Figure 2C, we expect C-Ada to crosslink efficiently with only DNA-2A and DNA-2T. In this scenario, the modified base C* in DNA-2G cannot enter the active site of C-Ada for effective crosslinking because it is held in the double

helical DNA structure through hydrogen bonding to the opposite base G.

Crosslinking Reactions between C-Ada and DNA

A random single-stranded oligonucleotide (DNA-1) with a thiol-tethered cytosine (C*) was synthesized and purified. The two-carbon thiol tether, protected with 2-aminoethanethiol mixed disulfide (Figure 3B), was introduced into the specific cytosine base C* by convertible nucleoside methodology [34–36]. Double-stranded oligonucleotides (DNA-2G, -2C, and -2T) were prepared by annealing DNA-1 with corresponding complementary strands. The 172 amino acid C-terminal fragment of Ada (C-Ada) was cloned, overexpressed, and purified. A mutant C-Ada with Cys139 mutated to Ala139 (C-Ada C139A) was also prepared and purified. The crosslinking reactions between C-Ada and the modified DNA were initiated by incubating 2.5 equivalents of DNA with 1 equivalent of protein. Control experiments with C-Ada C139A were performed in parallel.

Incubating 2.5 molar equivalents of single-stranded DNA-1 or double-stranded DNA-2G bearing a modified C* with C-Ada for 24 hr produced only the protein bands on the gel. No crosslinked complex was observed, as shown in Figure 3D (lanes 2 and 3). However, efficient crosslink formation between C-Ada and DNA-2A or DNA-2T was observed (~40% based on the molar equivalence of C-Ada). Control experiments with a mutant protein C-Ada C139A, in which the reactive Cys139 residue is mutated to an Ala residue, showed no crosslink formation between the protein and all four DNA probes (DNA-1, -2G, -2A, and -2T). This result indicates that the crosslinking reactions are specific for Cys139.

It is not surprising that C-Ada did not crosslink with single-stranded DNA-1. The protein has very low repair activity toward single-stranded DNA substrates [39]. Notably, C-Ada also failed to effectively crosslink with double-stranded DNA-2G. However, simply changing the base opposite of C* from G to a mismatched base A or T gave rise to an efficient formation of crosslinked products, as shown in Figure 3D. Clearly, in DNA-2A and DNA-2T, C* was inserted into the substrate binding pocket of C-Ada; and Cys139 attacked the disulfide tether to generate the desired crosslinked product. The inability of DNA-2G to crosslink with C-Ada indicates that C* in DNA-2G cannot access the active site pocket of C-Ada. With a thiol tether on the exocyclic 4 amino group, C* should still form normal Watson-Crick hydrogen bonds with base G on the complementary strand (C•G base pair is shown in Figure 4A). This base-pairing interaction holds C* in the double helix such that it cannot freely rotate out of the duplex structure. When mismatched bases are introduced opposite C* for DNA-2A and DNA-2T, a stable base pair no longer exists (Figure 4C) [37, 38]; therefore, C* can rotate out of the duplex structure becoming transiently extrahelical and be captured by C-Ada for efficient crosslinking. Our data suggests that C-Ada can detect the base lesion by capturing the damaged base that is already extrahelical, as shown in Figure 2C. C-Ada does not seem to search for lesions by flipping out every base as suggested in Figure 2A. Nor does it appear to locate lesions by directly detecting the adduct on the modified base.

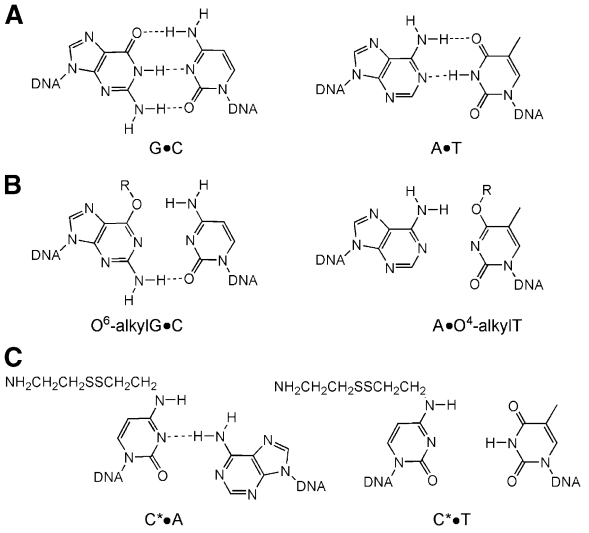

Figure 4. Base Pairs in DNA

(A) Watson-Crick base pairs of G•C and A•T.

(B) Alkylation damage on the O^6 position of guanine and O^4 position of thymine change the structure of the hydrogen-bonding interface of the base. The damaged base can no longer form a stable base pair with its opposite bases (C and A, respectively).

(C) Possible hydrogen-bonded structures of the C*•T and C*•A mismatches at neutral pH.

Normal G•C or A•T base pairs are stabilized by multiple hydrogen bonds as shown in Figure 4A. Addition of an alkyl group on the O^6 position of guanine or the O^4 position of thymine completely changes the hydrogen-bonding interface of these bases. The ability of these modified bases to form normal Watson-Crick base pairs is significantly weakened, as shown in Figure 4B. The damaged base can then rotate out of the duplex structure and be captured by an Ada protein migrating along the DNA (Figure 2C).

Besides the mechanism of simply capturing an extrahelical base, C-Ada may also facilitate the base-flipping step of a weakened base pair, thus playing some active role in the damage-searching process. Although C-Ada cannot disrupt normal stable Watson-Crick base pairs and directly detect a base adduct as we have shown; it could flip out the damaged base by recognizing the less stable nature of the lesioned base pairs (O^6-alkylG:C) relative to the normal Watson-Crick pairs. Thus, the protein can actively test every base pair and only flip out a damaged base in an unstable base pair. It could also distort DNA and destabilize the duplex structure to facilitate base flipping of a damaged base. We performed additional experiments to gain further insight into these mechanistic possibilities.

Time Course and Stability toward DTT of the Crosslinking Reaction

Next, we analyzed the time course of the crosslinking reactions with DNA-2A and DNA-2T. One equivalent of C-Ada was incubated with 2.5 equivalents of either DNA-2A or DNA-2T, and these reactions were quenched after 15, 30, 60, and 120 min by the addition of methyl methanethiolsulfonate (Figure 3E). Analysis of the time course samples showed that the crosslinking reaction between C-Ada and DNA-2A is fast and complete in 30–60 min in high efficiency (higher than 75%) as shown in Figure 3E. A large amount of crosslinked complex (~40%) was generated even after 15 min incubation (lane 1). The crosslinking reaction between C-Ada and DNA-2T is slower and less efficient. Crosslinked product was not observed on the gel until after 30 min incubation (lane 6). After 120 min of incubation, C-Ada crosslinked less with DNA-2T than DNA-2A. Both crosslinking reactions were modestly stable toward a treatment of strong reducing reagent dithiothreitol (DTT). In the presence of 10 equivalents of DTT, no significant effect on crosslinking reaction was detected. With the addition of more DTT, the amounts of disulfide crosslinked products were diminished (Figure 3F).

The kinetic and thermodynamic differences of the two crosslinking reactions can be rationalized by the difference in stability between the two mismatched base pairs. It is known that C•T is more stable than C•A [37]. Although addition of a thiol-tether on the N^4 position of cytosine will slightly decrease the helix stability due to electronic reasons [36], it is expected that the same trend in stability will hold with C*•T being more stable than C*•A. With A opposite C* in DNA-2A, Watson-Crick hydrogen bonds cannot form between the two bases because of steric repulsion between the two exocyclic amino groups. Instead, the two bases may form a weak hydrogen bond at neutral pH, as previously suggested (Figure 4C) [37]. With T opposite the modified base C* in DNA-2T, C* can stay inside the duplex structure either by base stacking with neighboring bases or by forming two weak hydrogen bonds with the opposite base T (Figure 4C) [38]. Thus, the C* in DNA-2A would be expected to adopt an extrahelical conformation more favorably than the C* in DNA-2T. If C-Ada simply captures bases that are already rotated out of the duplex DNA structure, it should crosslink with DNA-2A more effectively than with DNA-2T, as was observed.

We are intrigued by the idea that C-Ada may test every base pair and only selectively flip out a damaged base in an unstable base pair. In this scenario, C-Ada would possess a certain amount of activation energy that it could exert to disrupt unstable base pairs. This energy would not be enough to break normal Watson-Crick base pairs, but it is sufficient to disrupt unstable base pairs. If this activation energy is large, we would expect C-Ada to crosslink with DNA-2A and DNA-2T at a very similar rate. If this energy is very close to the stability of the two mismatched base pairs we used, we may observe very different kinetic behaviors of these crosslinking reactions. The stability of a mismatched C•T base pair versus a C•A base pair is around 0.2–0.5 kcal/mol [37], which corresponds to about 2- to 3-fold difference in reaction rate. Since the observed reaction rate difference of the crosslinking reactions involving the two DNA probes (DNA-2A and DNA-2T) seems to reflect the exact stability difference of the base pairs, we think the major role of the protein is merely capturing the extrahelical conformation of a weakly paired base, although the "actively testing" mechanism cannot be completely excluded.

Another possible mechanistic scenario is that C-Ada can distort DNA and destabilize the duplex structure to facilitate base flipping of a damaged base in an unstable base pair. To address this possibility, we analyzed the influence of the sequence surrounding the mismatched base pair on crosslinking reaction kinetics. Six more DNA probes were prepared as shown in Figure 3C. DNA-3G, DNA-3A, and DNA-3T all have an AT-rich sequence, and DNA-4G, DNA-4A, and DNA-4T were made with GC-rich sequence surrounding the modified cytosine. DNA-3G and DNA-4G, having stable C*•A base pair in the sequence, did not crosslink with C-Ada. DNA probes 3A, 3T, 4A, and 4T all crosslinked with C-Ada. Time course of the crosslinking reactions showed that the reaction rates of C-Ada with DNA-3A and DNA-4A, both having a C*•A base pair in the sequence, are similar to each other (Figures 3G and 3H). Reactions of C-Ada with DNA probes DNA-3T and DNA-4T, both having a C*•T base pair, did not show different kinetic behavior (Figures 3G and 3H). Together with crosslinking results obtained with DNA-2A and DNA-2T, the results seem to suggest that the crosslinking reaction rates between C-Ada and modified DNA is not significantly influenced by the overall stability of the duplex DNA. The reaction rate is highly dependent on the stability of the modified base pair. These results do not seem to support the mechanism that C-Ada distorts the duplex DNA to facilitate base flipping of a damaged base while the protein is scanning for base lesions.

Crosslink between hAGT and DNA

Although homologous to the *E. coli* C-Ada protein, human AGT has some very different properties. Human AGT can process larger adducts than C-Ada [21, 22] and shows significant activity toward repairing damages in single-stranded DNA [40]. It also removes alkyl adducts on the O^6 position of guanine when the alkylated guanine is mispaired with a thymine base [41]. The O^6-alkylguanine forms a reasonably stable base pair with a thymine base (see the hydrogen-bonding interfaces in Figure 4). Thus, the lesioned base is stabilized in an intrahelical conformation most of the time. In order to remove the alkyl group in this mismatched base pair, human AGT may need to actively locate and flip out the damaged base. The mechanism of this function could be different from that of C-Ada. Therefore, we investigated crosslinking reactions between human AGT and our DNA probes 1, 2G, 2A, and 2G.

A truncated construct of hAGT (1-179) containing the DNA repair active domains [21, 22] was cloned, overexpressed, and purified. A mutant hAGT with Cys145 mutated to Ala145 (hAGT C145A) was also prepared and purified. The crosslinking reactions between hAGT and modified DNA were conducted and assayed as described for C-Ada. Control experiments with hAGT

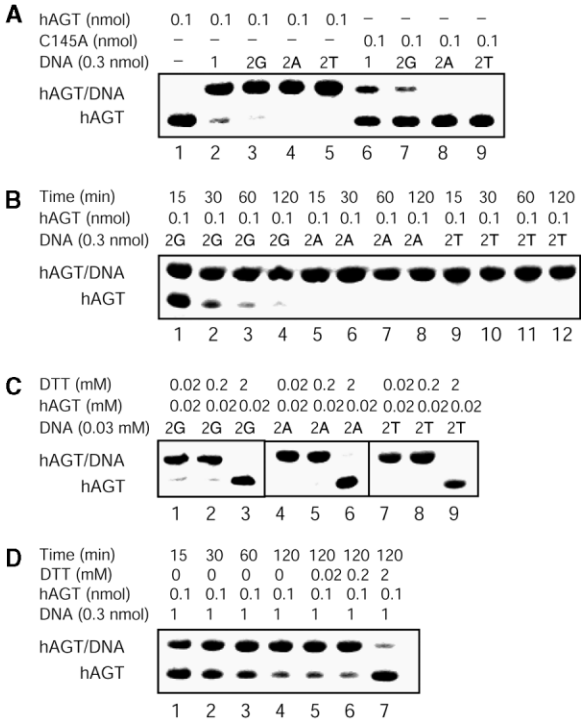

Figure 5. Disulfide Crosslinking between Human AGT and DNA-1, -2G, -2A, and -2T

All reactions were performed by incubating 10 μM of hAGT and 30 μM of DNA at 4°C.

(A) SDS gel analysis of the crosslinking reaction between hAGT and DNA-1, -2G, -2A, and -2T. Lane 1 is a size standard for hAGT. Lanes 2–5, crosslinking results between hAGT and DNA-1, -2G, -2A, and -2T after 24 hr incubation. Lanes 6–9, controls with hAGT C145A mutant.

(B) Time course of crosslinking reactions. Lanes 1–4, time course of crosslinking reaction between hAGT and DNA-2G. Lanes 5–8, time course of crosslinking reaction between hAGT and DNA-2A. Lanes 9–12, time course of crosslinking reaction between hAGT and DNA-2T.

(C) Effect of external thiol (DTT) on the reactions using varying amounts of DTT after 2 hr incubation. Lanes 1–3, crosslinking results between hAGT and DNA-2G. Lanes 4–6, crosslinking results between hAGT and DNA-2A. Lanes 7–9, crosslinking results between hAGT and DNA-2T.

(D) Time course and effect of external thiol (DTT) on the crosslinking reaction between hAGT and DNA-1. Lanes 1–4, time course of crosslinking reaction. Lanes 5–7, crosslinking results between hAGT and DNA-1 in the presence of varying amounts of DTT after 2 hr incubation.

C145A were performed in parallel. Human AGT crosslinked effectively with all four modified DNA (DNA-1, -2G, -2A, and -2T). Quantitative formation of crosslinked complex was observed when incubating 3 molar equivalents of the modified oligonucleotide DNA-2A or DNA-2T with 1 molar equivalent of hAGT for 24 hr (Figure 5A, lanes 4 and 5). Large amounts of crosslinked product (higher than 90% based on the molar equivalence of hAGT) also formed when single-stranded DNA-1 or double-stranded DNA-2G was used in the reaction (Figure 5A, lanes 2 and 3). Small amounts of uncrosslinked protein could be observed for both reactions.

The time course of crosslinking reactions between wild-type hAGT and all four DNA probes was examined. The crosslinking reactions between hAGT and DNA-2A or DNA-2T are remarkably fast, reaching completion after 15 min incubation (Figure 5B, lanes 5–12). The crosslinking reaction between hAGT and DNA-2G is slower. The reaction was completed in approximately 1 hr at 4°C with only a small amount of uncrosslinked protein remaining (Figure 5B, lanes 1–4). The crosslinking reaction between hAGT and single-stranded DNA-1 is the slowest, as indicted in Figure 5D (lanes 1–4). These results show that hAGT crosslinks with different DNA at various rates. All crosslinking reactions are modestly stable toward the treatment of a strong reducing reagent DTT (Figure 5C, lanes 1–9; Figure 5D, lanes 5–7).

Clearly, human AGT exhibited a different behavior toward crosslinking to DNA probes 1, 2G, 2A, and 2T in comparison to C-Ada. It even formed a crosslinked complex with single-stranded DNA-1. This observation is perhaps not so surprising given the fact that hAGT has been shown to repair damaged bases in single-stranded DNA. The protein must have evolved a method to check for base damage while binding to a single-stranded DNA, perhaps by checking every base in the single-stranded DNA. It remains to be seen how hAGT binds and recognizes damaged bases in both double-stranded and single-stranded DNA. In the light of a recent discovery that the alkylation DNA repair protein AlkB corrects damaged bases in RNA [9], it is not unreasonable to think that hAGT can also repair alkylated guanine or uracil in RNA.

What is unexpected to us is the observation that hAGT crosslinked with DNA-2G with high efficiency. Because of the stability of the C*•G base pair, the C* in DNA-2G is held intrahelically most of the time. C-Ada did not crosslink effectively with DNA-2G due to its inability to extrude C* from its intrahelical conformation. Crosslinking observed between hAGT and DNA-2G suggests that this protein may be able to actively search for damaged bases and extrude them from the duplex DNA structure. Either mechanism A or mechanism B shown in Figure 2 could be used by hAGT when it searches for the C* in DNA-2G.

Since the crosslinking reactions between hAGT and DNA-2A or DNA-2T proceed much faster than the reaction between hAGT and DNA-2G (Figure 5B), we suggest that hAGT has a mode to quickly capture unstable lesioned bases like C-Ada does. In a majority of the cases, guanine is alkylated while still base-pairing with C. This alkylated guanine, forming a weakened base pair with the opposite C, can be quickly captured by hAGT due to its unstable nature. This could be the mechanism used by hAGT to capture C* in DNA-2A or DNA-2T in the crosslinking reactions reported here. However, if the alkylated guanine is not repaired promptly, a thymine base can be incorporated opposite this damaged guanine base during DNA replication. The resulting O^6-alkylG•T base pair is fairly stable, and the damaged guanine is held inside the duplex DNA structure. The cellular mismatch repair system is used to correct this error in most cases [42, 43]. But it is also found that hAGT is involved to remove the alkyl group on the damaged guanine base opposite T [41]. In this case, hAGT may need to actively search and flip out the alkylated bases in a stable base

pair for repair. We think this mechanism is used by hAGT to locate C* in DNA-2G. So the primary mode to search for damage by hAGT is to simply migrate along DNA and recognize a lesioned base that is unstable in the DNA duplex. The base could sample the extrahelical conformation and be captured by hAGT or the protein could "sense" its unstable nature and actively flip it out. When there is no unstable lesion present, the protein may scan through DNA, detect potential intrahelical adducts, and actively extrude the damaged base from the duplex DNA structure for repair. This is a slower and less efficient mode compared to the detection of unstable base lesions. Whether the protein flips out every base for damage searching in this case is not clear.

Control experiments with the mutant protein hAGT C145A were conducted under the same conditions. A negligible amount of crosslinked product was generated in the reaction between the mutant protein hAGT C145A and either DNA-2A or DNA-2T (Figure 5A, lanes 8 and 9), indicating that the Cys145 residue of the wild-type protein is almost exclusively engaged in the crosslinking reaction. However, the formation of a significant amount of nonspecific crosslinking was observed when single-stranded DNA-1 was used (Figure 5A, lane 6). A small amount of nonspecific crosslink formation between the mutant protein and DNA-2G was detected as well (Figure 5A, lane 7).

The crystal structures of hAGT show that there are three exposed Cys residues, Cys5, Cys24, and Cys150, on the surface. Cys5 and Cys24 are both ligands to a structural zinc(II) ion. Cys150 is located on the proposed DNA binding surface of the protein [21, 22]. It is likely that this residue is engaged in DNA recognition and responsible for the nonspecific crosslinking between the hAGT C145A mutant protein and DNA-1 or DNA-2G. The extent of the nonspecific crosslinking is much less than the specific crosslinking with wild-type hAGT (Figure 5A). This indicates that the Cys145 residue is responsible for the formation of most crosslinked products in the reaction between wild-type hAGT and DNA-1 or DNA-2G. The reason for the lack of nonspecific crosslinking between hAGT C145A and DNA-2A or DNA-2T could be because C* in these DNA probes is not stable in the duplex DNA structure. The Cys residue responsible for the nonspecific crosslinking may not be able to access the disulfide moiety on an extrahelical cytosine base. However, we think that the lack of crosslinking between the hAGT C145A mutant protein with DNA-2A and DNA-2T is mainly due to quick binding of a C* in an unstable base pair in the active site of the mutant protein. The mutant protein can trap extrahelical C* very rapidly and bind the alkyl tether in its active site to form a stable specific complex. Thus, this C* could not crosslink to the surface exposed Cys residue, and the Cys145 residue would be shielded by the bound DNA from crosslinking to other modified DNA.

When C* is stabilized intrahelically as in DNA-2G, it must be flipped out by the protein, which results in a slower binding rate of C* to the protein. This slower specific binding rate allows for competitive nonspecific crosslinked products to be generated between the hAGT C145A mutant protein and DNA-2G. DNA-2G may also form a labile complex with the inactive C145A hAGT mutant protein because there is no unstable base pair present in DNA-2G for the protein to recognize and form a specific complex. Thus, nonspecific crosslinked products were produced after long time incubation. The specific binding of hAGT to the single-stranded DNA-1 is the weakest and slowest, and the most nonspecific crosslinked products were observed when the mutant protein was used.

In summary, by utilizing a chemical crosslinking technique, we suggest that C-Ada, a member of the O^6-alkylguanine DNA-alkyltransferase family, locates potential damaged bases by capturing the extrahelical lesions, rather than flipping out every base for damage detection. Although it was hypothesized that capturing an unstable lesioned base could be used in some systems [23], to our knowledge, it has not been previously confirmed by experiments. Our results provide a real example for this damage-searching mechanism. We also found that human AGT, a homologous protein of the *E. coli* C-Ada, can also efficiently capture bases in unstable base pairs. These results imply that detection of unstable lesions could be a general damage-searching mechanism in other base repair systems if the covalent damage destabilizes the normal Watson-Crick base pair. The unstable lesioned base could sample extrahelical conformation and be trapped by a repair protein like what we observed with C-Ada. Alternatively, the repair protein could test the stability of the damaged base pair and facilitate base flipping of the lesioned base. These modes provide efficient way for repair proteins to quickly locate potential base lesions.

We also showed that human AGT can actively detect and flip out damaged bases for repair when the lesioned base is stabilized in the duplex DNA structure. There are other base lesions that do not drastically alter the architecture and stability of duplex DNA, such as 8-oxoguanine opposite C, N^7-methylguanine opposite C, uracil opposite A, and O^6-alkylguanine opposite T. It was suggested that proteins that repair these lesions may need to extrude every base from the duplex DNA structure in order to locate damages [23]. It is possible that hAGT uses this type of damage-searching mode, but hAGT also possesses a very efficient way to identify the unstable nature of weakened base pairs. Other DNA repair systems could use similar strategies.

The approach we used here not only provides means to probe damage-searching mechanism in DNA repair systems, but it also enables us to trap covalently linked protein/DNA complexes for both C-Ada and hAGT. Disulfide crosslinking has proven to be very effective in stabilizing labile or nonspecific protein/DNA interactions for structural studies [31–33]. As we have shown here, C-Ada and hAGT can form an efficient and specific crosslink with DNA-2A or DNA-2T. This enables us to prepare homogenous protein/DNA complexes for further structural studies to understand how O^6-alkylguanine-DNA alkyltransferases work in bacteria and in humans.

Significance

How a base-flipping DNA repair protein locates its target base among thousands of base pairs in the

genome is a long-standing question. Several damage-searching mechanisms were proposed, but none of them has been confirmed experimentally. O^6-alkylguanine-DNA alkyltransferases have been proposed to repair the damaged base in an extrahelical manner like DNA glycosylase enzymes. Here we used a chemical disulfide crosslinking method to probe the damage-searching mechanism of the *E. coli* and human O^6-alkylguanine-DNA alkyltransferases. Our results suggest that *E. coli* C-Ada does not actively flip out every base for damage searching. It locates damaged bases by simply capturing a lesion that is sampling the extrahelical conformation transiently due to its significantly weakened base-pairing ability. The human protein, hAGT, also efficiently detects damaged bases that form unstable base pairs. But hAGT can extrude base lesions that are stabilized intrahelically in duplex DNA structure in a less efficient process. We believe capturing an extrahelical lesioned base and detecting unstable nature of a damaged base pair are simple, efficient, and general ways to locate damaged bases for base repair proteins. Actively flipping a base out for damage searching may not be necessary in many cases. Our methods can be utilized to study other DNA repair and DNA modification systems. The crosslinking presented here enables covalently linked protein/DNA complexes to be prepared for both *E. coli* C-Ada and human AGT in large yields, which can be used for further structural characterization.

Experimental Procedures

Construction, Expression, and Purification of C-Ada

Ada construct used in this study is the C-terminal Glu185-Arg354 fragment of the *E. coli ada* gene. It was cloned into PET30a vector (Novagen, Madison, WI) between NdeI and HindIII restriction sites. The sequence was numbered Met1-Arg172, with Glu3-Arg172 corresponding to Glu185-Arg354 in the full-length Ada. The expression plasmid was transformed into *E. coli* BL21 (DE3) cells. The cells were grown aerobically with kanamycin (30 μg/ml) at 37°C until the OD_{600} was 0.7. IPTG (1 mM) was added and the cells were grown for an additional 4 hr at 30°C. All subsequent steps were performed at 4°C. The cells were harvested by centrifugation and stored at −80°C. The cell pellet was resuspended in 30 ml of lysis buffer (100 mM Tris-HCl [pH 7.34], 300 mM NaCl, 5% glycerol, 2 mM $CaCl_2$, 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol), disintegrated by sonication, and centrifuged at 12,000 rpm for 20 min. The supernatant was then added to 3 equivalents of buffer A (10 mM Tris-HCl [pH 7.34]), loaded onto S Sepharose cation exchange column (Amersham Biosciences, Piscataway, NJ) that had been equilibrated with buffer A, and eluted with a linear gradient of NaCl (0.0–1.0 M). The fractions containing the protein were concentrated by ultrafiltration (Centricon YM10 membrane; Amicon, Millipore Corporation, Bedford, MA) and purified further with a Mono-S cation exchange column (Amersham Biosciences) using a linear gradient of NaCl (0.0–1.0 M). Point mutation in C-Ada (C139A) was introduced by megaprimer mutagenesis. The sequences of C-Ada and mutant C-Ada were confirmed by sequencing of the entire coding sequence.

Construction, Expression, and Purification of Human AGT

An expression vector pGEX-6P-2 (Amersham Biosciences) containing a GST-tagged truncated form of human AGT protein (1-179) was a gift from Dr. G.L. Verdine. The *hAGT* gene was cloned between the BamHI and EcoRI restriction sites with a PreScission Protease recognition sequence between the GST domain and hAGT domain of the fusion protein. Point mutation in hAGT (C145A) was introduced by megaprimer mutagenesis. The sequences of wild-type and mutant hAGT were confirmed by sequencing of the entire coding sequence.

Expression plasmids were transformed into *E. coli* BL21 (DE3) cells. The growth and lysis of the cells were performed as described with C-Ada. The supernatant was loaded onto GSTrap FF column (2 × 1 ml; Amersham Biosciences) and washed with lysis buffer. The fusion protein was eluted with 5 ml of elution buffer (10 mM Tris-HCl [pH 7.34], 100 mM NaCl, 10 mM 2-mercaptoethanol, and 10 mM reduced glutathione). The fractions containing the fusion protein was treated with PreScission Protease overnight at 4°C to cleave the GST-tag. The resulting hAGT protein was purified further with the Mono-S cation exchange column (Amersham Biosciences).

Synthetic Oligonucleotides

Oligodeoxynucleotides were synthesized on an Applied Biosystems 392 DNA synthesizer. The thiol-tethered oligonucleotide was prepared by incorporation of O^4-triazolyl-dU-CE phosphoramidite (Glen Research) at the modified positions during solid-phase synthesis. The diamine disulfide tether was added via a postsynthetic modification/deprotection method described previously [34–36]. All oligonucleotides were purified by denaturing polyacrylamide gel electrophoresis. Concentrations of the oligonucleotides were estimated by UV at 260 nm.

Crosslinking Reaction and Analysis

Purified proteins were dialyzed into a buffer containing 10 mM Tris-HCl (pH 7.4) and 100 mM NaCl. The same buffer was used for the crosslinking reactions. Typically the protein (20 μM for C-Ada and 10 μM for hAGT) and the thiol-tethered duplex DNA (50 μM for C-Ada and 30 μM for hAGT) were incubated at 4°C for varying periods of time in the crosslinking buffer (20–100 μl). The reaction was then quenched by the addition of a thiol-capping reagent, methyl methanethiolsulfonate, to a final concentration of 20 mM for 10 min at room temperature. After SDS-PAGE, loading dye free of reducing agents was added, and the samples were analyzed by SDS-PAGE.

Acknowledgments

This research was supported by the University of Chicago. The plasmids encoding full-length *E. coli ada* and truncated human AGT are gifts from Dr. Gregory L. Verdine.

Received: April 23, 2003
Revised: June 24, 2003
Accepted: July 21, 2003
Published: September 19, 2003

References

1. Wood, R.D., Mitchell, M., Sgouros, J., and Lindahl, T. (2001). Human DNA repair genes. Science *291*, 1284–1289.
2. Friedberg, E.C., Walker, G.C., and Siede, W. (1995). DNA Repair and Mutagenesis (Washington, DC: ASM Press).
3. Labahn, J., Scharer, O.D., Long, A., Ezaz-Nikpay, K., Verdine, G.L., and Ellenberger, T.E. (1996). Structural basis for the excision repair of alkylation-damadged DNA. Cell *86*, 321–329.
4. Yamagata, Y., Kato, M., Odawara, K., Tokuno, Y., Nakashima, Y., Matsushima, N., Yasumura, K., Tomita, K., Ihara, K., Fujii, Y., et al. (1996). Three-dimensional structure of a DNA repair enzyme 3-methyladenine DNA glycosylase II, from *Escherichia coli*. Cell *86*, 311–319.
5. Bruner, S.D., Norman, D.P.G., and Verdine, G.L. (2000). Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. Nature *403*, 859–866.
6. Trewick, S., Henshaw, T.F., Hausinger, R.P., Lindahl, T., and Sedgwick, B. (2002). Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA damage. Nature *419*, 174–178.
7. Falnes, P.O., Johansen, R.F., and Seeberg, E. (2002). AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. Nature *419*, 178–182.
8. Duncan, T., Trewick, S.C., Koivisto, P., Bates, P.A., Lindahl, T., and Sedgwick, B. (2002). Reversal of DNA alkylation damage

by two human dioxygenases. Proc. Natl. Acad. Sci. USA *99*, 16660–16665.
9. Arnes Aas, P., Otterlei, M., Falnes, P.O., Vågbo, C.B., Skorpen, F., Akari, M., Sundheim, O., Bjørås, M., Slupphaug, G., Seeberg, E., et al. (2003). Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. Nature *421*, 859–863.
10. Yarosh, D.B. (1985). The role of O^6-methylguanine-DNA methyltransferase in cell survivial, mutagenesis, and carcinogenesis. Mutat. Res. *145*, 1–16.
11. Lindahl, T., Sedgewick, B., Sekiguchi, M., and Nakabeppu, Y. (1988). Regulation and expression of the adaptive response to alkylating agents. Annu. Rev. Biochem. *57*, 133–157.
12. Demple, B. (1990). Self-methylation by suicide DNA repair enzymes. In Protein Methylation, W.K. Pair and S. Kim, eds. (Boca Raton, FL: CRC Press), pp. 285–304.
13. Pegg, A.E. (1990). Mammalian O^6-alkylguanine-DNA alkyltransferase: regulation and importance in response to alkylating carcinogenic and therapeutic agents. Cancer Res. *50*, 6119–6129.
14. Mitra, S., and Kaina, B. (1993). Regulation of repair of alkylation damage in mammalian genomes. Prog. Nucleic Acid Res. Mol. Biol. *44*, 109–142.
15. Keppler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H., and Johnsson, K. (2003). A general method for the covalent labeling of fusion proteins with small molecules *in vivo*. Nat. Biotechnol. *21*, 86–89.
16. Zak, P., Kleibl, K., and Laval, F. (1994). Repair of O^6-methylguanine and O^4-methylthimine by the human and rat O^6-methylguanine-DNA methyltransferases. J. Biol. Chem. *269*, 730–733.
17. Brent, T.P., Houghton, P.J., and Houghton, J.A. (1985). O^6-alkylguanine-DNA alkyltransferase activity correlates with the therapeutic response of human rhabdomyosarcoma xenografts to 1-(2-chloroethyl)-3-(trans-4-methylcyclohexyl)-1-nitrosourea. Proc. Natl. Acad. Sci. USA *82*, 2985–2989.
18. Mattern, J., Eichhorn, U., Kaina, B., and Volm, M. (1998). O^6-methylguanine-DNA methyltransferase activity and sensitivity to cyclophosphamide and cisplatin in human lung tumor xenografts. Int. J. Cancer *77*, 919–922.
19. Kaina, B., Fritz, G., Mitra, S., and Coquerelle, T. (1991). Transfection and expression of human O^6-methylguanine-DNA methyltransferase (MGMT) cDNA in Chinese hamster cells: the role of MGMT in protection against the genotoxic effects of alkylating agents. Carcinogenesis *12*, 1857–1867.
20. Moore, M.H., Gulbis, J.M., Dodson, E., Demple, B., and Moody, P.C. (1994). Crystal structure of a suicidal DNA repair protein: the Ada O^6-methylguanine-DNA methyltransferase from *E. coli*. EMBO J. *13*, 1495–1501.
21. Daniels, D.S., Mol, C.D., Arvai, A.S., Kanugula, S., Pegg, A.E., and Tainer, J.A. (2000). Active and alkylated human AGT structures: a novel zinc site, inhibitor and extrahelical base binding. EMBO J. *19*, 1719–1730.
22. Wibley, J.E.A., Pegg, A.E., and Moody, P.C.E. (2000). Crystal structure of the human O^6-alkylguanine-DNA alkyltransferase. Nucleic Acids Res. *28*, 393–401.
23. Verdine, G.L., and Bruner, S.D. (1997). How do DNA repair proteins locate damaged bases in the genome? Chem. Biol. *4*, 329–334.
24. Verdine, G.L. (1994). The flip side of DNA methylation. Cell *76*, 197–200.
25. Roberts, R., and Cheng, X. (1998). Base flipping. Annu. Rev. Biochem. *67*, 181–198.
26. Hornby, D.P., and Ford, G.C. (1998). Protein-mediated base flipping. Curr. Opin. Biotechnol. *9*, 354–358.
27. Klimasauskas, S., Kumar, S., Roberts, S.J., and Cheng, X. (1994). HhaI methyltransferase flips its target base out of the DNA helix. Cell *76*, 357–369.
28. Reinisch, K.M., Chen, L., Verdine, G.L., and Lipscomb, W.N. (1995). The crystal structure of HaeIII methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. Cell *82*, 143–153.
29. Slupphaug, G., Mol, C.D., Kavil, B., Arvai, A.S., Krokan, H.E., and Tainer, J.A. (1996). A nucleotide-flipping mechanism from the structure of human uracil DNA glycosylase bound to DNA. Nature *384*, 87–92.
30. Demple, B., Sedgwick, B., Robin, P., Totty, N., Waterfield, M.D., and Lindahl, T. (1985). Active site and complete sequence of the suicidal methyltransferase that counters alkylation mutagenesis. Proc. Natl. Acad. Sci. USA *82*, 2688–2692.
31. Huang, H., Chopra, R., Verdine, G.L., and Harrison, S.C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase at 2.7 Å resolution: implications of comformational changes for polymerization and inhibition mechanism. Science *282*, 1669–1675.
32. Huang, H., Harrison, S.C., and Verdine, G.L. (2000). Trapping of a catalytic HIV reverse transcriptase-template:primer complex through a disulfide bond. Chem. Biol. *7*, 355–364.
33. He, C., and Verdine, G.L. (2002). Trapping distinct structural states of a protein/DNA interaction through disulfide crosslinking. Chem. Biol. *9*, 1297–1303.
34. Xu, Y.Z., Zheng, Q., and Swann, P.F. (1992). Synthesis of DNA containing modified bases by postsynthetic substitution. Synthesis of oligomers containing 4-substituted thymine: O^4-alkylthymine, 5-methylcytosine, N^4-(dimethylamino)-5-methylcytosine and 4-thiothymine. J. Org. Chem. *57*, 3839–3845.
35. MacMillan, A.M., and Verdine, G.L. (1990). Synthesis of functionally tethered oligonucleotides by the convertible nucleoside approach. J. Org. Chem. *55*, 5931–5933.
36. MacMillan, A.M., and Verdine, G.L. (1991). Engineering tethered DNA molecules by the convertible nucleoside approach. Tetrahedron *14*, 2603–2616.
37. Allawi, H.T., and SantaLucia, J., Jr. (1998). Nearest-neighbor thermodynamics of internal A:C mismatches in DNA: sequence dependence and pH effects. Biochemistry *37*, 9435–9444.
38. Allawi, H.T., and SantaLucia, J., Jr. (1998). Thermodynamics of internal C:T mismatches in DNA. Nucleic Acids Res. *26*, 2694–2701.
39. Lindahl, T., Demple, B., and Robins, P. (1982). Suicidal inactivation of the *E. coli* O^6-methylguanine-DNA methyltransferase. EMBO J. *1*, 1359–1363.
40. Luu, K.X., Kanugula, S., Pegg, A.E., Pauly, G.T., and Moschel, R.C. (2002). Repair of oligodeoxyribonucleotides by O^6-alkylguanine-DNA alkyltransferase. Biochemistry *41*, 8689–8697.
41. Lips, J., and Kaina, B. (2001). Repair of O^6-methylguanine is not affected by thymine base pairing and the presence of MMR proteins. Mutat. Res. *487*, 59–66.
42. Ullah, S., and Day, R.S., III. (1992). Incision at O^6-methylguanine:thymine mispairs in DNA by extracts of human cells. Biochemistry *31*, 7998–8008.
43. Lari, S.-U., Al-Khodairy, F., and Paterson, M.C. (2002). Substrate specificity and sequence preference of G:T mismatch repair: incision at G:T, O^6-methylguanine:T and G:U mispairs in DNA by human Cell Extracts. Biochemistry *41*, 9248–9255.