doi:10.1016/j.jmb.2005.05.028 J. Mol. Biol. (2005) 350, 657–666

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

The Structure of the Human AGT Protein Bound to DNA and its Implications for Damage Detection

Erica M. Duguid[1], Phoebe A. Rice[2] and Chuan He[1]*

[1]*Department of Chemistry 5735 South Ellis Avenue The University of Chicago Chicago, IL 60637, USA*

[2]*Department of Biochemistry and Molecular Biology, 920 East 58th Street, Chicago, IL 60637 USA*

**Corresponding author*

O^6-Alklyguanine-DNA alkyltransferase (AGT) is an important DNA repair protein that protects cells from mutagenesis and toxicity arising from alkylating agents. We present an X-ray crystal structure of the wild-type human protein (hAGT) bound to double-stranded DNA with a chemically modified cytosine base. The protein binds at two different sites: one at the modified base, and the other across a sticky-ended DNA junction. The protein molecule that binds the modified cytosine base flips the base and recognizes it in its active site. The one that binds ends of neighboring DNA molecules partially flips an overhanging thymine base. This base is not inserted into the active-site pocket of the protein. These two different hAGT/DNA interactions observed in the structure suggest that hAGT may not detect DNA lesions by searching for the adduct itself, but rather for weakened and/or distorted base-pairs caused by base damage in the duplex DNA. We propose that hAGT imposes a strain on the DNA duplex and searches for DNA regions where the native structure is destabilized. The structure provides implications for pyrimidine recognition, improved inhibitor design, and a possible protein/protein interaction patch on hAGT.

Keywords: DNA repair; human AGT; damage detection

Introduction

DNA is subject to chemical modification by various environmental factors from sunlight to water and by internally generated metabolites. Prime targets for modification are the heterocylic bases.[1,2] Because of this, multiple DNA repair proteins have evolved to repair damage that, if unchecked, may lead to mutations. These repair proteins work in a variety of ways, depending on the type and location of the damage. Direct repair, typically requiring only one protein, is often employed to repair alkylation damage. This method is used by the N-terminal domain of Ada in *Escherichia coli* to remove methyl groups from methylated backbones,[3,4] and by the AlkB proteins, which use a complexed iron atom to oxidatively demethylate adducts on N^1-adenine and N^3-cytosine.[5–7]

Another of the direct repair proteins is O^6-alkylguanine-DNA alkyltransferase (AGT, EC 2.1.1.63) which removes numerous alkyl groups from the O6 position of guanine, preferentially in double-stranded DNA.[3,4,8–10] The lesion, which can lead to G:C→A:T point mutations, is highly mutagenic if not repaired.[11–14] This protein family is widely conserved;[15,16] the best studied being the C-terminal domain of Ada in *E. coli* and hAGT, known also as MGMT, in humans. An active-site cysteine residue in the conserved -(I/V)PCHR(V/I)- motif of the protein transfers the alkyl group from the damaged guanine base to itself through an S_N2 mechanism (Figure 1(a)). O^6-Alkylthymine serves as a substrate for the AGT proteins, including hAGT, *in vitro*;[17] however, whether this base lesion is actually repaired inside human cells is still unclear.[18,19] hAGT is adept at removing a wide range of alkyl groups from the canonical methyl group to larger adducts.[20–22] The alkylation of the protein is irreversible; after one repair reaction the protein is non-functional and is subsequently degraded by the cell. Since many cancer chemotherapeutic agents such as carmustine 1,3-bis(2-chloroethyl-1-nitrosourea (BCNU), procarbazine, and temozolomide act by modifying the O6 position of guanine and inducing apoptosis, hAGT has been of interest to the medical community.[23] The presence of hAGT results in

Abbreviations used: AGT, O^6-alkylguanine-DNA alkyltransferase; hAGT, human AGT.

E-mail address of the corresponding author: chuanhe@uchicago.edu

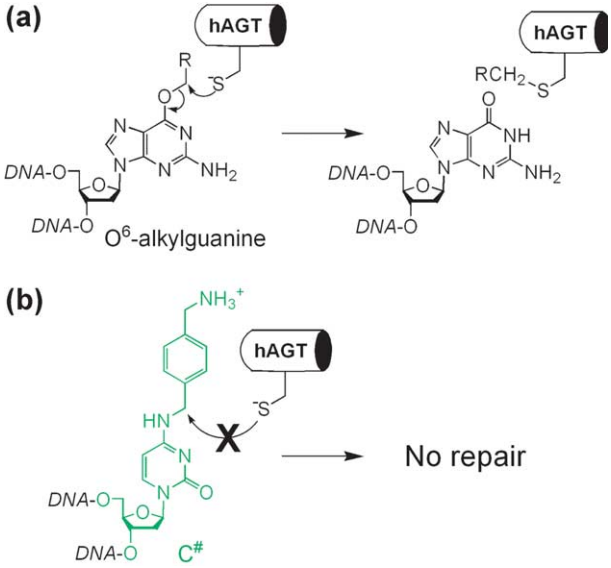

Figure 1. (a) Suicide repair of O^6-alkylguanine in DNA by the hAGT protein. (b) With a C–N linkage, the *p*-xylylenediamine group of the modified N^4-alkylcytosine ($C^{\#}$) cannot be displaced by hAGT.

resistance of cancer cells to the chemotherapies mentioned above and efforts are underway to find efficient suicide substrates to inactivate the protein.[24] The small molecule O^6-benzylguanine, which benzylates the active-site Cys residue, is currently being investigated as a co-chemotherapeutic agent but is limited by low affinity and poor water solubility.[25,26]

It was established recently that hAGT is a metalloprotein binding a zinc(II) ion near the N terminus with residues Cys5, Cys24, His29, and His85.[27] This zinc ion, though located far from the active site, increases the reaction rate of the wild-type hAGT by ~60-fold over metal-stripped hAGT,[28] and seems to aid in lowering the pK_a of the active-site residue Cys145.[29] This effect is postulated to be due to increased stability and proper alignment of the N and C-terminal domains. Crystal structures of hAGT with and without zinc have been obtained.[27,30] No significant change of the protein conformation near the active site was observed. Thus, the role of the zinc ion is unclear.

One of the most interesting questions surrounding DNA repair proteins is how they search through the immense space of the chromosome to identify and subsequently repair a small number of lesions.[31] These proteins, which as expected are sequence-independent, must balance between binding to DNA tightly enough to recognize and perform the repair and retaining sufficient freedom to move through the genome. The active site of many base-specific DNA-repair proteins, including hAGT, is buried inside the protein, which necessitates the flipping of the damaged base out of the DNA helical structure for repair.[32–42] Direct detection of the damage adduct seems unlikely for hAGT, given the fact that it can process different lesions. Flipping out each base sequentially may not be the most efficient method. Our previous experiments indicate that a mispaired aberrant base in the helical structure is more likely to be detected by AGT than the same base in a Watson–Crick base-pair.[43] We suggest that finding weakened base-pairs or distortion induced by base damage in the DNA structure may be an efficient way for the AGT proteins to locate possible damage sites, since the primary target base lesion of AGT, O^6-alkylguanine, has an altered Watson–Crick surface that misaligns with that of cytosine.

Very recently, structures of an inactive hAGT mutant bound to an O^6-methylguanine-containing duplex DNA and of the wild-type hAGT cross-linked to the same sequence of DNA containing a mechanistic inhibitor were solved at 3.2 Å resolution.[44] Unexpectedly, the structure revealed that hAGT binds the minor groove of the duplex DNA. We have used a different strategy to stabilize hAGT/DNA complexes. Here, we present the X-ray structure of the wild-type hAGT bound to DNA in two different contexts: one with the protein recognizing a chemically modified base and the other with the protein bound at a DNA junction flipping an unmodified base. A modified N^4-alkylcytosine base ($C^{\#}$) that cannot be repaired by the protein is used to stabilize one of the hAGT/DNA interactions (Figure 1(b)). Several new features are shown in this structure as compared to the previous structure: the zinc(II) atom, which is missing from the previous hAGT/DNA structure, is present. A pyrimidine base is recognized in the active site in one complex. Most interestingly, one of the hAGT molecules binds the ends of two neighboring duplex DNAs and partially inserts an overhanging thymine into its active-site pocket. The structure highlights the importance of DNA perturbation as the target for damage-searching by hAGT as well as molecular basis for the diverse substrate tolerances of hAGT.

Results and Discussion

Design of oligonucleotides that form specific protein/DNA complexes with hAGT

A chemically modified base was used to achieve the crystallization of the hAGT/DNA complex in spite of the rather low binding affinity of the protein for DNA; $K_d = 0.8$ μM for hAGT binding to native DNA and only two- to threefold tighter to DNA containing O^6-methylguanine.[45] The modified nucleotide is cytosine with the organic group *p*-xylylenediamine attached to the N^4 position ($C^{\#}$ in Figure 1(b)). This base was designed with several advantages in mind: (i) the nitrogen–carbon bond between the cytosine and the adduct cannot be severed by Cys145 in the wild-type protein; (ii) the aromatic ring in the tail may interact with the

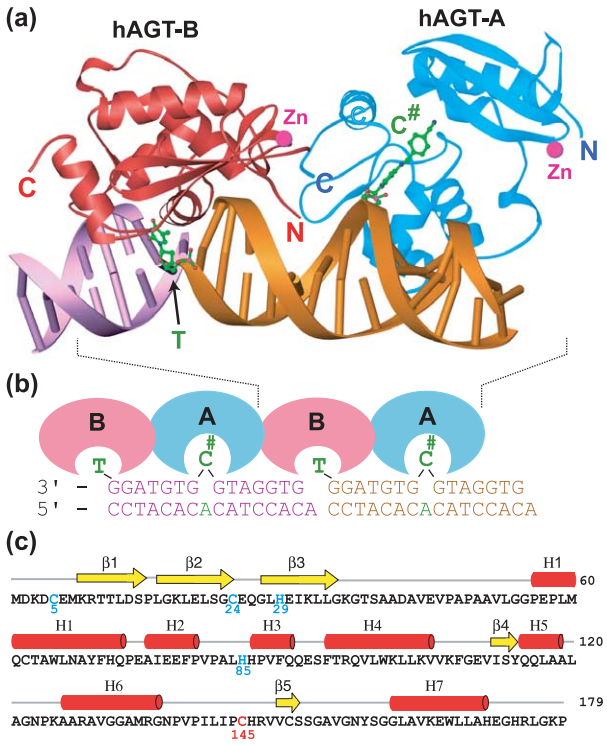

Figure 2. Structure of the hAGT/DNA complex. (a) Two different complex forms are in the crystal structure: hAGT-A flips $C^{\#}$ and hAGT-B recognizes a terminal overhanging T (note that this thymine base is inserted into the pocket of hAGT-B only partially). A structural zinc(II) is present in both protein molecules. (b) Schematic of the hAGT/DNA structure showing DNA sequence and overall binding. (c) Sequence and secondary structure of truncated hAGT used in this study, with helices depicted as red cylinders and β-strands as yellow arrows. The four Cys residues that coordinate zinc are colored blue and the active-site Cys145 is colored red.

hydrophobic substrate recognition pocket of hAGT, as revealed by previous structures of the protein;[27] (iii) analysis of the benzylated hAGT structure[27] shows the presence of hydrogen bond-accepting groups near the active-site pocket that may interact favorably with the primary amine group at the end of our adduct. Our previous work indicates that hAGT binds preferentially to mismatched base-pairs.[43] On the basis of this observation, we placed an adenine base opposite $C^{\#}$ to further stabilize a specific hAGT/DNA complex.

Overall structure

A truncated hAGT missing C-terminal residues 180–207 was used in the structural study. Retention of full methyl transferase activity had been shown for a more extreme truncation (1–175).[27] Crystals of this protein bound to a 16-mer duplex of DNA containing A/T overhangs (Figure 2(b)) were obtained. $\langle I/\sigma I \rangle$ falls below 2 at 3.4 Å, better than 3 and 4 along a^*, b^* and c^*, respectively. The structure was determined by molecular replacement and refined to 3.0 Å (Figure 2(a)). The electron-density map is continuous for the DNA and for most of the protein from residues 5–174. However, of the three monomers in each asymmetric unit, two contain a break between residues 35 and 46, while the third has breaks between residues 35 and 51, and between residues 150 and 157. The stretch from residues 35–51 in hAGT seems to be quite flexible; some residues in this area were missing in the recently solved hAGT/DNA complex structure.[44] The overall fold of hAGT is very similar to previously determined structures.[27,30,44] However, the composition and conformation of the bound DNA regions are different in the new structure.

Two different hAGT/DNA interactions present in the structure

The DNA duplexes are *B*-form and stack end to end at the overhangs as expected, with two distinct and anti-parallel pseudo-continuous helices running through each asymmetric unit. On each helix, the modified base is flipped out of the DNA helix and into the active site of an hAGT monomer (Figures 2 and 3(a)). However, on one of the helices another hAGT monomer binds at the junction of the DNA fragments. This monomer also displaces a base from the duplex DNA but this base, the overhanging thymine, is inserted into the active-site pocket only partially (Figures 2 and 3(b)). The monomers binding to the modified base and junction of the same DNA helix are referred to as hAGT-A and hAGT-B, respectively, and the monomer binding to the second helix of DNA is called hAGT-C. As the conformations of A and C are very similar, most of the discussion that follows will compare A and B.

The protein has two domains, as revealed previously.[27,30] The active site is hosted in the C-terminal domain, which interacts with DNA directly. Our structure confirms the recent observation of DNA minor groove binding by the helix(H5)-turn-helix(H6) motif near the C terminus (Figures 2 and 4).[44] However, unlike that structure,

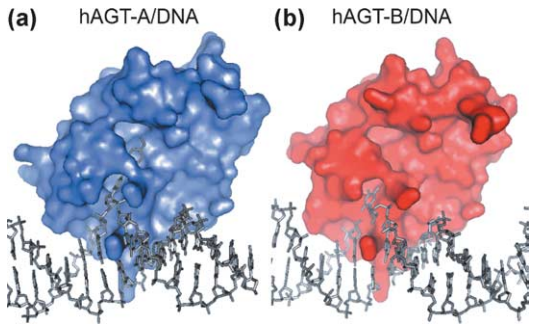

Figure 3. Comparison of the overall protein/DNA interactions between (a) hAGT-A and (b) hAGT-B.

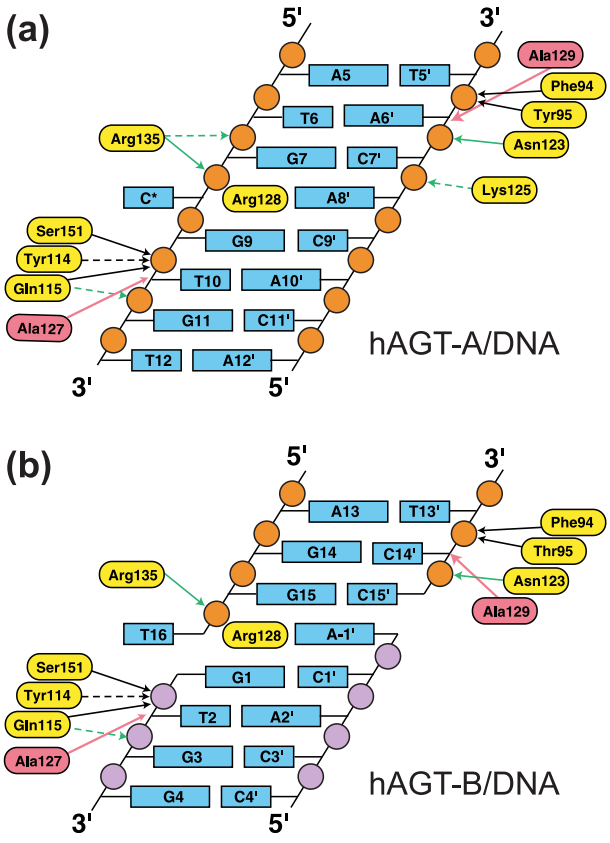

Figure 4. Schematic of hAGT/DNA interactions. Arrows indicate direct (continuous) or possible water-mediated (broken) interaction from main chain (black) or side-chain (green) of protein residues. Red arrows show contacts between two Ala side-chains and DNA backbones that help widen the minor groove. (a) Interactions between hAGT-A and DNA region containing $C^{\#}$. (b) Interactions between hAGT-B and the DNA junction containing the overhanging T base.

ours contains a zinc(II) ion bound by Cys5, Cys24, H29 and H55 near the N terminus (Figure 2). This zinc atom is ~20 Å away from the active Cys145 residue and does not alter the active-site geometry in any obvious way.

Each protein monomer induces a slight bend of the DNA over four base-pairs away from the protein, a total 30° roll angle for monomers A and C and a 15° roll angle for B when analyzed with the program 3DNA.[46] The observed DNA bending by hAGT-B is similar to that of the complex structures reported recently;[44] hAGT-A and hAGT-B appear to impose more bending on DNA. The exact reason is unclear to us. It could be due to different crystal packing, or different DNA binding when purine or pyrimidine is recognized in the active site of the protein (hAGT-B does not insert the thymine base into its active site completely). More structural work may provide further insights.

The protein monomers bind the DNA through backbone amides, side-chains and/or water-mediated interactions to only a handful of phosphate groups spreading over a 7 bp region (Figure 4). Residues that bind to the strand with the flipped base are located along the first helix (H5, Figure 2(c)) of the helix-turn-helix motif and around the small β sheet in the C-terminal domain (β5, Figure 2(c)). Contacts between the complementary strand and the protein occur only on three phosphate groups 3′ of the modification.

Recognition of $C^{\#}$ by hAGT-A

In the hAGT-A/DNA interaction, Arg128, the "arginine finger" necessary for repair of base damage in duplex DNA,[27] replaces the flipped-out base in the DNA duplex to fill the pocket vacated by the base and to hydrogen bond to the opposite adenine base (Figures 5(a) and 6). A water molecule

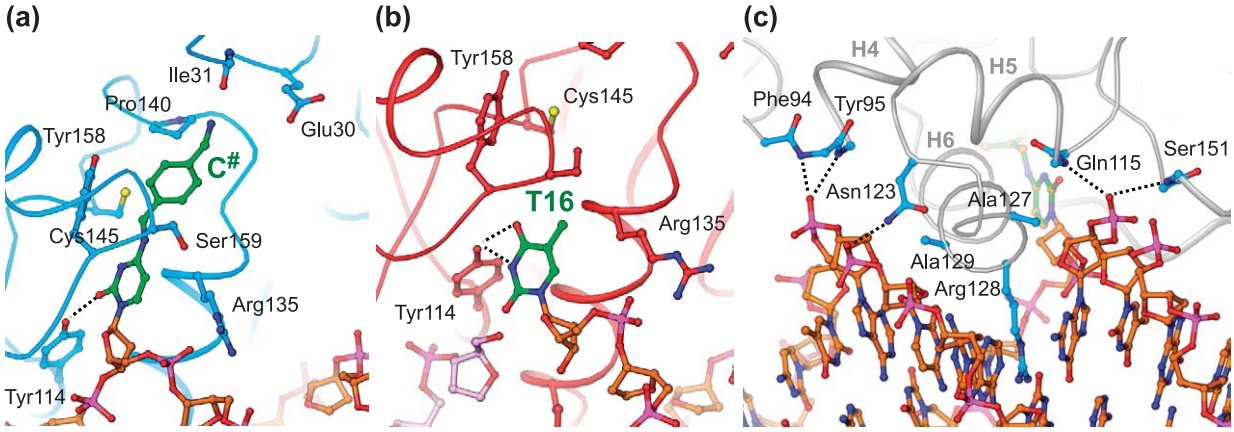

Figure 5. Protein/DNA interactions observed in two hAGT monomers. (a) Recognition of $C^{\#}$ by hAGT-A. Tyr114 forms a strong hydrogen bond (2.5 Å) with the O2 atom of the base. Cys145 is in an in-line position to attack the carbon atom linked to the N4 atom of $C^{\#}$. (b) The overhang base thymine is partially inserted into hAGT-B (compare the relative position of the base with $C^{\#}$ in hAGT-A). Tyr114 is within hydrogen bonding distance to T. (c) The protein makes three major contacts to the DNA (hAGT-A is shown here). The second helix (H6) of the HTH is buried in the DNA minor groove.

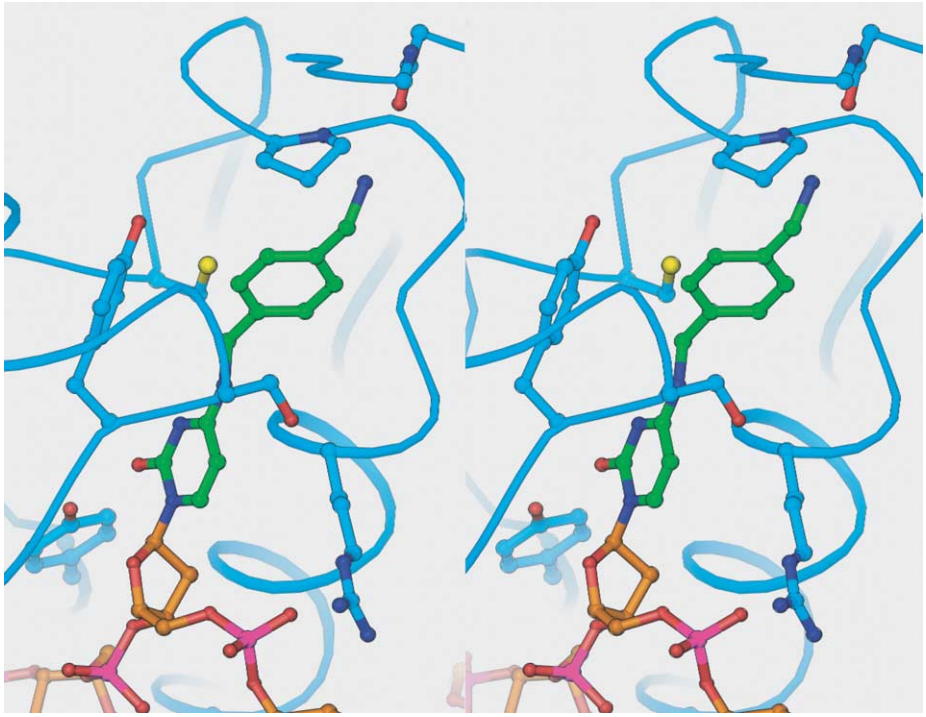

Figure 6. A wall-eyed stereo view of the active site of hAGT-A.

fills the gap left by the sugar and forms hydrogen bonds bridging Arg128 and the DNA backbone. C$^{\#}$ sits, as expected, in the active-site pocket between Pro140, Tyr158, and Ser159 (Figures 3(a) and 5(a)). While there is no density for the amine tail it could be modeled close to the backbone oxygen atom of Ile30.

The binding of C$^{\#}$ by hAGT-A reveals pyrimidine recognition by hAGT. It is known that isolated hAGT can remove alkyl groups from the O4 position of a thymine base.[17] Our structure shows that Tyr114 forms a hydrogen bond to O2 of the cytosine, implicating the presence of the same interaction if a damaged thymine was present (Figure 5(a)). This Tyr residue has been shown as a key residue in recognizing the O^6-alkylguanine lesion by hydrogen bonding to the N3 atom of the modified base.[44] This hydrogen bonding interaction helps protonate N3 of a damaged guanine base to accelerate removal of the alkyl adduct.[27] For efficient repair of O^4-alkylthymine, the protonation should occur on the N3 atom of an alkylated thymine base. Tyr114 cannot play this role, since it forms a hydrogen bond with the O2 atom of a pyrimidine base and is too far from N3 (Figures 5(a) and 6). This might be one of the reasons that O^4-alkylthymine is repaired less efficiently than O^6-alyklguanine by hAGT.[17] However, Tyr114 may still be able to protonate N3 through the isomerization of a protonated O2 by the phenol side-chain. This function will facilitate removal of the alkyl group at the O4 position of a thymine base.

The use of C$^{\#}$ as the substrate mimic allows crystallization of the wild-type hAGT with DNA, since Cys145 cannot remove the alkyl group attached on the N4 position of C$^{\#}$. The sulfur atom of this Cys residue is located 3.45 Å away from the carbon atom linked to the N4 atom of C$^{\#}$. Cys145 is in an in-line position to displace the alkyl adduct on C$^{\#}$ (Figures 5(a) and 6). Cys145 is placed so that it is consistent with a water-mediated hydrogen bond network with His146 and Glu172 as observed in the structures without bound DNA.[27,30] This hydrogen bond network has been proposed to help remove the proton on the sulfur atom of Cys145 to facilitate a nucleophilic attack on the alkyl lesion.

Flipping of an undamaged thymine base by hAGT-B

One of the most interesting aspects of the structure is the presence of hAGT-B, which binds the junction of two neighboring DNA molecules and flips out an overhanging thymine base (Figures 2, 3(b) and 4(b)). The extrahelical thymine is not fully inserted into the active site and is far from the active Cys145 residue (Figures 2(b) and 3(b)). The N3 and O4 atoms of the thymine base are within hydrogen bonding distance of the phenol oxygen atom of Tyr114 (Figure 5(b)). The 5-methyl group of the thymine base packs with the aliphatic portion of the Arg135 side-chain. Arg128 in hAGT-B remains intercalated in the base stack but is ~1 Å farther away from the orphaned adenine base than its counterpart in hAGT-A; a strong hydrogen bonding interaction to the orphaned base does not appear to be important for binding.

Protein/DNA interactions and implications for the base-flipping mechanism

Globally, hAGT makes three major contacts with DNA, one to each phosphate backbone and a third along the groove (Figures 4 and 5(c)). In the first interaction a phosphate group on the strand containing the flipped base is bracketed by three backbone amide groups at positions 114, 115 and 151. The interaction involving the backbone of Gln115 is strengthened by its placement at the positive end of helix H5′s dipole. At the second point of contact, a phosphate group on the complementary strand interacts with the dipole of H4 through the backbone of Thr95, as well as hydrogen bonding with the amide nitrogen atom. The side-chain of residue Asn123 participates in the interaction to the same DNA backbone. In the third interaction, the second helix (H6) of the HTH resides deep in the widened minor groove. This helix is rich with small, hydrophobic residues such as Ala and Gly. These residues pack against the floor of the minor groove and the flanking phosphodiester backbones. The side-chains of Ala127 and Ala129, pointing towards sugars in the opposite direction, help define the extent of the minor groove widening (Figure 5(c)). The residue between these two Ala is the arginine finger Arg128 that penetrates into the DNA helical structure. All together, these three major contacts impose strain to bend the DNA and widen the minor groove. As the result, the DNA minor groove bound by hAGT-A is widened by >3 Å relative to perfect *B*-DNA, as observed in the recent structure with O^6-methylguanine-containing DNA recognized by hAGT.[44]

Clearly, the junction of the neighboring DNAs is more flexible, and therefore offers a favorable site to relax the strain created by hAGT binding. The same strain imposed on DNA by hAGT-A is used to bend DNA, which may contribute to base flipping of $C^{\#}$. Base flipping seems to be a direct result of the protein binding to the less rigid DNA structure. Recognition of the flipped base may not provide the predominant driving force for the formation of the complex. This notion seems to be reinforced by the fact that the thymine base is inserted into hAGT-B only partially (Figure 3(b)). We propose from these observations that the base-flipping hAGT protein preferentially searches for DNA regions where the native structure is already destabilized. An alternative explanation for the observed partial flipping of the thymine base could be the existence of a "gate-keeping" mechanism for hAGT. As demonstrated recently for hOGG1, the base-flipping DNA repair proteins may possess a mechanism that prefers the specific lesion to be inserted fully into the active site, while undamaged bases are more likely to be at a second binding conformation.[47] In the absence of an alkyl lesion, a normal base such as thymine may not be recognized favorably in the hAGT's active site, but instead may be bound in a second gate-keeping site of the protein.

Flexible recognition of base lesions

A noticeable difference between hAGT-A and hAGT-B lies in the final α-helix (H7) and the loop region (150–160) that links this helix to the rest of the C-terminal domain. This loop region contains residues (Ser159 and Tyr158) that contribute to recognition of the flipped base in hAGT-A. The helix and the loop move further away from the core of the protein in hAGT-A than in hAGT-B (Figure 2(c)). The different placement of this loop and terminal helix between hAGT-A and hAGT-B reflects the flexibility of this region and unusual ability of hAGT to recognize lesions varying greatly in size and shape.

Burial of hydrophobic surfaces has been postulated to be the cause of the increased binding and repair rate for O^6-benzylguanine both in DNA and as a free base over O^6-methylguanine. As expected, the aromatic xylene ring of $C^{\#}$ is packed against Pro140, Tyr158 and Ser159 in the structure (Figure 4(a)). In addition, the amine tail of $C^{\#}$ is close enough to either the Glu30 side-chain or the Ile31 carbonyl oxygen atom to form a hydrogen bond, depending on the rotation (Figure 4(a)).

The zinc(II) site and a potential protein/protein interaction patch

A zinc(II) ion bound by Cys5, Cys24, H29 and H55 is present in our hAGT structure (Figure 7). As mentioned above, this zinc ion in hAGT appears to play some role in the rate of reaction with damaged bases. Comparison of this zinc-bound hAGT/DNA structure with the hAGT/DNA structures without zinc shows little change around the active site. However, structures with the metal ion have a better-ordered N-terminal region than those without zinc. In particular a greater ordering of the amino acid residues in the stretch from 36 to 51 is observed. This area contains proline at positions 47,

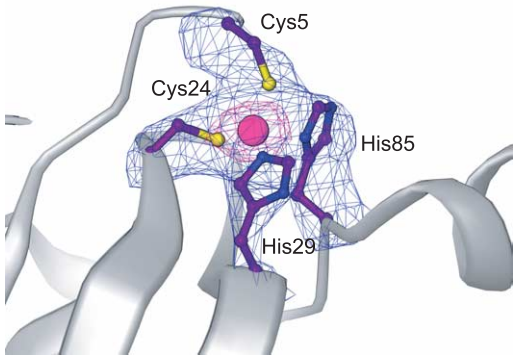

Figure 7. Close-up view of the tetrahedral zinc(II) site coordination in hAGT-B. The metal is shown as a pink sphere, and the side-chains of the ligand residues are blue, with sulfur atoms colored yellow. The map represents simulated annealing omitting the Zn^{2+} and the four ligand residues ($F_{obs}-F_{calc}$, the blue electron density surface is at 2.6σ and the pink is at 7σ).

49, 56, and 58, as well as methionine nearby at 7 and 60 (Figure 2(c)). This proline-rich area, missing from the previously published hAGT/DNA complex,[44] curves over the top of the β-sheet toward the zinc site in monomers A and B. This stretch is also rich with hydrophobic residues such as Val, Ala and Ile (Figure 2(c)). This patch makes significant protein/protein contact in crystal packing in these monomers.

Analysis of protein/protein interaction surfaces has suggested proline[48] and methionine[49] as residues that play predominant roles. It has been hypothesized for some time that hAGT may interact with other proteins in order to be removed from the DNA and degraded after repair. Zinc may help order this hydrophobic surface for protein/protein recognition. The ordered hydrophobic residues may help mediate self-assembly to form protein complexes in lesion repair, as suggested previously.[50] The role of the zinc ion could be to stabilize the overall structure of hAGT and thus achieve a higher repair activity, as suggested previously.[28] These hypotheses will be tested in future studies.

Implications from the structure

Base-flipping DNA repair proteins may use different mechanisms to search for lesions. Base modification may give rise to a weakened base-pair. The modified base in this base-pair can sample the extrahelical conformation more frequently, which contributes to the capture of the base by a repair protein, as suggested by us and by others.[43,51–53] Additionally, compression of backbone and kinking DNA has been suggested as a potential damage-detection mechanism for glycosylases.[54] The present structure provides further insights into this intriguing problem. This structure reveals the binding of the wild-type hAGT proteins to two different regions of a duplex DNA helix. The first molecule, hAGT-A, recognizes a mismatched base-pair and flips out a modified cytosine base as designed. The second molecule, hAGT-B, binds the junction of two neighboring DNAs and partially flips an overhanging thymine base. Analysis of protein/DNA interactions in these two structures suggests that hAGT imposes a strain upon binding to DNA. A weakened base-pair in DNA or the junction of two DNA duplexes can relieve this strain and thus are recognized preferentially by hAGT. We propose that this repair protein searches for DNA regions where the native structure is destabilized through base modification. Base flipping is a result of a specific binding of the protein to a DNA sequence that contains a weakened and/or non-Watson–Crick base-pair. The same strategy may be adopted by other base repair proteins.

It should be noted that the partial flipping of the thymine base in hAGT-B could be due to a potential gate-keeping mechanism of hAGT, as proposed for hOGG1 recently by Verdine and co-workers.[47] This base, at the end of the duplex DNA structure, is prone to sampling the extrahelical conformation, which might be captured readily by hAGT-B. We are in the process of stabilizing hAGT on duplex DNA containing no modification. Further structural studies on the discrete "non-specific" hAGT/DNA complexes should lead to more insights into the damage-searching mechanisms of these proteins.

Materials and Methods

Construction, expression, and purification of human AGT

An expression vector pGEX-6P-2 (Amersham Biosciences) containing a glutathione-*S*-transferase-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 the EcoRI restriction sites with a PreScission protease recognition sequence between the glutathione-*S*-transferase domain and hAGT domain of the fusion protein. The sequence was confirmed by sequencing of the entire coding sequence. Expression and purification of this hAGT construct have been described.[43]

Synthetic oligonucleotides

The modified oligodeoxynucleotide was synthesized on an Applied Biosystems 392 DNA synthesizer by incorporating an O^4-triazolyl-dU-CE phosphoramidite (Glen Research) at the position to be modified during solid-phase synthesis. The *p*-xylylenediamine tail was added by a postsynthetic modification/deprotection/cleavage step,[55–57] during which the resin was heated to 65 °C for 16 hours in an aqueous 15% (v/v) acetonitrile solution saturated with *p*-xylylenediamine purchased from Sigma. The modified oligocucleotide 5′-GTGGATGC$^{\#}$GTGTAGGT-3′ was purified by urea/polyacrylamide gel electrophoresis and annealed with the complementary strand 5′-CCTACACACATCCACA. Concentrations of the oligonucleotides were estimated by measuring absorbance at 260 nm.

Crystallization, structure determination, and refinement

The annealed DNA and hAGT (0.8 mM) were premixed in 10 mM Tris–HCl, 100 mM NaCl, 5 mM 2-mercaptoethanol in a protein to DNA molar ratio of 1:1.2. The hanging-drop, vapor-diffusion crystallization condition consisted of 1 μl of protein/DNA solution mixed with 1 μl of well solution (100 mM citrate buffer (pH 5.7), 22% (w/v) PEG 4000, 100 mM NaCl, 50 mM $MgCl_2$) equilibrated over 0.8 ml of well solution and grown at 4 °C.

The crystals were soaked briefly in well solution with 20% (v/v) glycerol before freezing. Data were collected at beamline 14-BM-C of BioCARS at the Advanced Photon Source at Argonne National Laboratory. The crystals belong to the monoclinic space group $P2_1$ with unit cell dimensions $a=58.15$ Å, $b=102.71$ Å, $c=87.94$ Å, $\beta=106.77°$ All data were integrated and scaled with the HKL2000 1.97 suite of programs.[58] Phases were calculated from a molecular replacement solution identified using the previously determined hAGT/DNA complex,[44] with EPMR 2.5. Model building and refinement were

Table 1. X-ray data collection, structure determination and refinement statistics

Wavelength (Å)	0.9000
Resolution (Å)	50–3.4 (along a^*), better than 3 (b^*), and 4 (c^*)
R_{sym}[a]	0.110
Total no. observations	69,838
No. unique observations	16,496
Completeness	83.0
Space group	$P2_1$
No. non-hydrogen atoms	4948
No. water molecules	6
$R_{working}$[b] (%)	0.246
R_{free} (%)	0.285
r.m.s. deviation	
Bond lengths (Å)	0.010
Bond angles (deg.)	1.4

[a] R_{sym} is the unweighted R value on I of symmetry-related reflections $\left(\sum_{hkl} |I_{hkl} - \langle I_{hkl}\rangle| / \sum_{hkl} I_{hkl}\right)$.
[b] $R_{working} = \sum_{hkl} |F_{obshkl} - F_{calchkl}| / \sum_{hkl} |F_{obshkl}|$, R_{free} is calculated from 5% of reflections not used in model refinement.

conducted with CNS 1.1 and CCP4 5.0.2. Refinement used energy minimization plus a combination of segmental 2-fold and 3-fold non-crystallographic symmetry restraints. Data collection and refinement statistics are shown in Table 1. The structure has two duplex DNA fragments and three monomers of hAGT per asymmetric unit and a solvent content of 66% (v/v).

Protein Data Bank accession number

The coordinates have been deposited in the RCSB Protein Data Bank under accession number 1YFH.

Acknowledgements

We thank the staff of the BioCARS sector at Argonne National Laboratories Advanced Photon Source for their help in data collection, Dr J. Piccirilli for the use of his DNA synthesizer, Dr N.-S. Li for assistance in DNA synthesis, Dr G. L. Verdine for the hAGT plasmid, and the reviewers for insightful comments. This research was supported by the University of Chicago, the Searle Scholars Program (to C.H.), the G&P Foundation for Cancer Research (to C.H.), the W. M. Keck Foundation Distinguished Young Scholars Program (to C.H.), and a National Institutes of Health (NIH) grant for Predoctorial Training at the Interface of Chemistry and Biology (to E.M.D.). Use of the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science, under contract no. W-31-109-Eng-38. Use of the BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources, under grant number RR07707.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2005.05.028

References

1. Friedberg, E. C., Walker, G. C. & Siede, W. (1995). *DNA Repair and Mutagenesis*, ASM Press, Washington, DC.
2. Wood, R. D., Mitchell, M., Sgouros, J. & Lindahl, T. (2001). Human DNA repair genes. *Science*, **291**, 1284–1289.
3. Lindahl, T., Sedgewick, B., Sekiguchi, M. & Nakabeppu, Y. (1988). Regulation and expression of the adaptive response to alkylating agents. *Annu. Rev. Biochem.* **57**, 133–157.
4. Sedgwick, B., Robins, P., Totty, N. & Lindahl, T. (1988). Functional domains and methyl acceptor sites of the *Escherichia coli* Ada protein. *J. Biol. Chem.* **263**, 4430–4433.
5. Falnes, P. Ø., Johansen, R. F. & Seeberg, E. (2002). AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature*, **419**, 178–182.
6. Trewick, S., Henshaw, T. F., Hausinger, R. P., Lindahl, T. & Sedgwick, B. (2002). Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA damage. *Nature*, **419**, 174–178.
7. Sedgwick, B. (2004). Repairing DNA-methylation damage. *Nature Rev. Mol. Cell Biol.* **5**, 148–157.
8. Yarosh, D. B. (1985). The role of O^6-methylguanine-DNA methyltransferase in cell survivial, mutagenesis, and carcinogenesis. *Mutat. Res.* **145**, 1–16.
9. 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.
10. Mitra, S. & Kaina, B. (1993). Regulation of repair of alkylation damage in mammalian genomes. *Prog. Nucl. Acid Res. Mol. Biol.* **44**, 109–142.
11. Pegg, A. E. (2000). Repair of O^6-alkylguanine by alkyltransferases. *Mutat. Res.* **462**, 83–100.
12. Kyrtopoulos, S. A., Anderson, L. M., Chhabra, S. K., Souliotis, V. L., Pletsa, V., Valavanis, C. & Georgiadis, P. (1997). DNA adducts and the mechanism of carcinogenesis and cytotoxicity of methylating agents of environmental and clinical significance. *Cancer Detect. Prev.* **21**, 391–405.
13. Kyrtopoulos, S. A. (1998). DNA adducts in humans after exposure to methylating agents. *Mutat. Res.* **405**, 135–143.
14. Margison, G. P. & Santibanez-Koref, M. F. (2002). O^6-alkylguanine-DNA alkyltransferase: role in carcinogenesis and chemotherapy. *BioEssays*, **24**, 255–266.

15. Kleibl, K. (2002). Molecular mechanisms of adaptive response to alkylating agents in *Escherichia coli* and some remarks on O^6-methylguanine DNA-methyltransferase in other organisms. *Mutat. Res.* **512**, 67–84.
16. Eisen, J. A. & Hanawalt, P. C. (1999). A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat. Res.* **435**, 171–213.
17. Zak, P., Kleibl, K. & 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.
18. Richardson, F. C., Dyroff, M. C., Boucheron, J. A. & Swenberg, J. A. (1985). Differential repair of O^4-alkylthymine following exposure to methylating and ethylating hepatocarcinogens. *Carcinogenesis*, **6**, 625–629.
19. Singer, B. (1986). O-Alkyl pyrimidines in mutagenesis and carcinogenesis: occurrence and significance. *Cancer Res.* **46**, 4879–4885.
20. Elder, R. H., Margison, G. P. & Rafferty, J. A. (1994). Differential inactivation of mammalian and *Escherichia coli* O^6-alkylguanine-DNA alkyltransferases by O^6-benzylguanine. *Biochem. J.* **298**, 231–235.
21. Goodtzova, K., Kanugula, S., Edara, S., Pauly, G. T., Moschel, R. C. & Pegg, A. E. (1997). Repair of O^6-benzylguanine by the *Escherichia coli* Ada and Ogt and the human O^6-alkylguanine-DNA alkyltransferases. *J. Biol. Chem.* **272**, 8332–8339.
22. Pegg, A. E., Boosalis, M., Samson, L., Moschel, R. C., Byers, T. L., Swenn, K. & Dolan, M. E. (1993). Mechanism of inactivation of human O^6-alkylguanine-DNA alkyltransferase by O^6-benzylguanine. *Biochemistry*, **32**, 11998–12006.
23. Gerson, S. L. (2002). Clinical relevance of MGMT in the treatment of cancer. *J. Clin. Oncol.* **20**, 2388–2399.
24. Dolan, M. E. & Pegg, A. E. (1997). O^6-Benzylguanine and its role in chemotherapy. *Clin. Cancer Res.* **3**, 837–847.
25. Pegg, A. E., Kanugula, S., Edara, S., Pauly, G. T., Moschel, R. C. & Goodtzova, K. (1998). Reaction of O^6-benzylguanine-resistant mutants of human O^6-alkylguanine-DNA alkyltransferase with O^6-benzylguanine in oligodeoxyribonucleotides. *J. Biol. Chem.* **273**, 10863–10867.
26. Dolan, M. E., Chae, M. Y., Pegg, A. E., Mullen, J. H., Friedman, H. S. & Moschel, R. C. (1994). Metabolism of O^6-benzylguanine, an inactivator of O^6-alkylguanine-DNA alkyltransferase. *Cancer Res.* **54**, 5123–5130.
27. Daniels, D. S., Mol, C. D., Arvai, A. S., Kanugula, S., Pegg, A. E. & Tainer, J. A. (2000). Active and alkylated human AGT structures: a novel zinc site, inhibitor and extrahelical base binding. *EMBO J.* **19**, 1719–1730.
28. Rasimas, J. J., Kanugula, S., Dalessio, P. M., Ropson, I. J., Fried, M. G. & Pegg, A. E. (2003). Effects of zinc occupancy on human O^6-alkylguanine-DNA alkyltransferase. *Biochemistry*, **42**, 980–990.
29. Guengerich, P. F., Fang, Q., Liu, L., Hachey, D. L. & Pegg, A. E. (2003). O^6-Alkylguanine-DNA Alkyltransferase: low pK_a and high reactivity of cysteine 145. *Biochemistry*, **42**, 10965–10970.
30. Wibley, J. E. A., Pegg, A. E. & Moody, P. C. E. (2000). Crystal structure of the human O^6-alkylguanine-DNA alkyltransferase. *Nucl. Acids Res.* **28**, 393–401.
31. Verdine, G. L. & Bruner, S. D. (1997). How do DNA repair proteins locate damaged bases in the genome? *Chem. Biol.* **4**, 329–334.
32. Verdine, G. L. (1994). The flip side of DNA methylation. *Cell*, **76**, 197–200.
33. Klimasauskas, S., Kumar, S., Roberts, S. J. & Cheng, X. (1994). *Hha*I methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357–369.
34. Reinisch, K. M., Chen, L., Verdine, G. L. & Lipscomb, W. N. (1995). The crystal structure of *Hae*III methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base-pairing. *Cell*, **82**, 143–153.
35. Slupphaug, G., Mol, C. D., Kavil, B., Arvai, A. S., Krokan, H. E. & Tainer, J. A. (1996). A nucleotide-flipping mechanism from the structure of human uracil DNA glycosylase bound to DNA. *Nature*, **384**, 87–92.
36. Labahn, J., Scharer, O. D., Long, A., Ezaz-Nikpay, K., Verdine, G. L. & Ellenberger, T. E. (1996). Structurral basis for the excision repair of alkylation-damaged DNA. *Cell*, **86**, 321–329.
37. Yamagata, Y., Fujii, S., Kato, M., Odawara, K., Tokuno, Y., Nakashima, Y. *et al*. (1996). Three-dimensional structure of a DNA repair enzyme 3-methyladenine DNA glycosylase II, from *Escherichia coli*. *Cell*, **86**, 311–319.
38. Roberts, R. & Cheng, X. (1998). Base flipping. *Annu. Rev. Biochem.* **67**, 181–198.
39. Hornby, D. P. & Ford, G. C. (1998). Protein-mediated base flipping. *Curr. Opin. Biotechnol.* **9**, 354–358.
40. Bruner, S. D., Norman, D. P. G. & Verdine, G. L. (2000). Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature*, **403**, 859–866.
41. Fromme, J. C., Banerjee, A., Huang, S. J. & Verdine, G. L. (2003). Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nature*, **427**, 652–656.
42. David, S. S. & Williams, S. D. (1998). Chemistry of glycosylases and endonucleases involved in base-excision repair. *Chem. Rev.* **98**, 1221–1261.
43. Duguid, E. M., Mishina, Y. & He, C. (2003). How do DNA repair proteins locate potential base lesions? A chemical crosslinking method to investigate the damage-searching mechanism of O^6-alkylguanine-DNA alkyltransferases. *Chem. Biol.* **10**, 827–835.
44. Daniels, D. S., Woo, T. T., Luu, K. X., Noll, D. M., Clarke, N. D., Pegg, A. E. & Tainer, J. A. (2004). DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nature Struct. Biol. Mol. Biol.* **11**, 714–720.
45. Meyer, A. S., McCain, M. D., Fang, Q., Pegg, A. E. & Spratt, T. E. (2003). O6-alkylguanine-DNA alkyltransferases repair O^6-methylguanine in DNA with Michaelis–Menten-like kinetics. *Chem. Res. Toxicol.* **16**, 1405–1409.
46. Lu, X.-J. & Olsen, W. K. (2003). 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucl. Acids Res.* **31**, 5108–5121.
47. Banerjee, A., Yang, W., Karplus, M. & Verdine, G. L. (2005). Structure of a repair enzyme interrogating undamaged DNA elucidated recognition of damaged DNA. *Nature*, **434**, 612–618.
48. Kay, B. K., Williamson, M. P. & Sudol, M. (2000). The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J.* **14**, 231–241.
49. Jones, S. & Thorton, J. M. (1997). Prediction of protein–protein interaction sites using patch analysis. *J. Mol. Biol.* **272**, 121–132.
50. Rasimas, J. J., Pegg, A. E. & Fried, M. G. (2003). DNA-binding mechanism of O6-alkylguanine-DNA

alkyltransferase. Effects of protein and DNA alkylation on complex stability. *J. Biol. Chem.* **278**, 7973–7980.
51. Cao, C., Jiang, Y. L., Stivers, J. T. & Song, F. (2004). Dynamic openning of DNA during the enzymatic search for a damaged base. *Nature Struct. Biol. Mol. Biol.* **11**, 1230–1236.
52. Krosky, D. J., Schwarz, F. P. & Stivers, J. T. (2004). Linear free engery correlation for enzyme base flipping: how do damaged base-pairs faciliate specific recognition? *Biochemistry*, **43**, 4188–4195.
53. Mishina, Y. & He, C. (2003). Probing the structure and function of the *Escherichia coli* DNA alkylation repair AlkB protein through chemical cross-linking. *J. Am. Chem. Soc.* **125**, 8730–8731.
54. Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E. & Tainer, J. A. (1998). Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J.* **17**, 5214–5226.
55. MacMillan, A. M. & Verdine, G. L. (1990). Synthesis of functionally tethered oligonucleotides by the convertible nucleoside approach. *J. Org. Chem.* **55**, 5931–5933.
56. MacMillan, A. M. & Verdine, G. L. (1991). Engineering tethered DNA molecules by the convertible nucleoside approach. *Tetrahedron*, **14**, 2603–2616.
57. Xu, Y. Z., Zheng, Q. & Swann, P. F. (1992). Synthesis of DNA containing modified bases by postsynthetic substitution. Synthesis of oligomers containing 4-substituted thymine: O^4-alkylthyine, 5-methyl-cytosine, N^4-(dimethylamino)-5-methylcytosine and 4-thiothymine. *J. Org. Chem.* **57**, 3839–3845.
58. Otwinowski, Z. M. W. (1996). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.

Edited by P. J. Hagerman

(Received 8 February 2005; received in revised form 4 May 2005; accepted 14 May 2005)