Cell, Vol. 86, 321–329, July 26, 1996, Copyright ©1996 by Cell Press

# Structural Basis for the Excision Repair of Alkylation-Damaged DNA

**Jörg Labahn,* Orlando D. Schärer,† Alexander Long,* Khosro Ezaz-Nikpay,† Gregory L. Verdine,† and Tom E. Ellenberger***
*Harvard Medical School
Department of Biological Chemistry and Molecular Pharmacology
Boston, Massachusetts 02115
†Harvard University
Department of Chemistry and Chemical Biology
Cambridge, Massachusetts 02138

## Summary

**Base-excision DNA repair proteins that target alkylation damage act on a variety of seemingly dissimilar adducts, yet fail to recognize other closely related lesions. The 1.8 Å crystal structure of the monofunctional DNA glycosylase AlkA (E. coli 3-methyladenine-DNA glycosylase II) reveals a large hydrophobic cleft unusually rich in aromatic residues. An Asp residue projecting into this cleft is essential for catalysis, and it governs binding specificity for mechanism-based inhibitors. We propose that AlkA recognizes electron-deficient methylated bases through π–donor/acceptor interactions involving the electron-rich aromatic cleft. Remarkably, AlkA is similar in fold and active site location to the bifunctional glycosylase/lyase endonuclease III, suggesting the two may employ fundamentally related mechanisms for base excision.**

## Introduction

The role of DNA as a stable repository for genetic information is constantly challenged by the chemical reactivity of the DNA bases (Lindahl, 1993). Nearly all atoms of the bases are subject to some form of hydrolysis, oxidative damage, or alkylation. Base modifications can arise either spontaneously or in response to exogeneous chemical reagents and, if left uncorrected, these lesions may cause mutations or impair DNA replication (Friedberg et al., 1995). For example, the cellular methyl donor S-adenosylmethionine reacts nonenzymatically with DNA to produce 7-methylguanine ($m^7G$) and 3-methyladenine ($m^3A$) (Rydberg and Lindahl, 1982). Both of these lesions can interfere with DNA-binding by regulatory proteins, and $m^3A$ also poses a potentially lethal blockage of DNA synthesis. $m^7G$ and $m^3A$ are selectively excised from DNA by monofunctional DNA N-glycosylases that cleave the C1′-N glycosylic bond, releasing the damaged bases from DNA (Figure 1) (Sakumi and Sekiguchi, 1990; Friedberg et al., 1995). The resulting abasic site (AP-site) is then repaired by excision/replacement DNA synthesis.

Escherichia coli encodes two N-alkylpurine DNA glycosylases, the constitutively expressed Tag protein ($m^3A$ DNA glycosylase I), and the inducible AlkA protein ($m^3A$ DNA glycosylase II) (Thomas et al., 1982). Whereas Tag excises only $m^3A$, AlkA efficiently removes $m^3A$, $m^7G$, and a number of minor adducts (Figure 1) including those produced by several clinically useful alkylating agents (Mattes et al., 1996). The broad substrate specificity of AlkA mirrors that of mammalian N-alkylpurine DNA glycosylases, yet these enzymes share little primary sequence homology (Seeberg et al., 1995). A long-standing mystery has been how AlkA selectively excises such chemically diverse alkylation products having little in common except a positive charge and the presence of an aberrant methyl group.

The specific targeting of base excision-repair enzymes to discrete lesions embedded in a vast excess of normal duplex DNA poses a formidable challenge, and in most cases the structural basis for the selectivity is not known. However, recent crystal structures of uracil DNA glycosylases suggest that the target base rotates out of the DNA helix and inserts into a pocket at the enzyme active site (Mol et al., 1995; Savva et al., 1995). Selectivity for deoxyuridine results from the steric and hydrogen-bonding complementarity of the cognate base and the enzyme active site. The cocrystal structure of the thymine dimer-specific glycosylase T4 endonuclease V (Endo V protein) shows that the enzyme sharply kinks DNA at the position of the thymine dimer, but makes no direct contacts to the thymine base targeted for cleavage (Vassylyev et al., 1995). Instead, the adenine complementary to the 5′-thymine of the dimer is flipped out of the DNA helix and bound into a hydrophobic cavity in the protein.

We report the crystal structure of the E. coli AlkA protein at 1.8 Å resolution. The surface of the AlkA protein has a prominent cleft lined with aromatic side chains that could provide a binding pocket for an electron-deficient alkylated base. The catalytic activity of AlkA is abolished by the substitution of an aspartic acid within the pocket (Asp-238) to asparagine, yet this mutant retains high affinity for DNA-based inhibitors, thus providing evidence that the aromatic cavity is the enzyme active site. The two domains of AlkA bordering its aromatic pocket bear strong resemblence to E. coli endonuclease III (Endo III) (Kuo et al., 1992; Thayer et al., 1995), a DNA glycosylase with additional endonuclease (AP-lyase) activity that processes oxidized pyrimidines in DNA. The corresponding cleft of endonuclease III is decorated with polar and charged residues, including an aspartic acid at the same position as the catalytically important Asp-238 of AlkA. AlkA thus appears to use a novel mechanism for molecular recognition that exploits π-electronic complementarity between its electron-rich active site cleft and electron-deficient bases. The close structural correspondence of AlkA and Endo III suggests an underlying similarity in the catalytic mechanisms of glycosylases and glycosylase/lyases.

## Results and Discussion

### Overall Structure of the AlkA Protein

The crystal structure of AlkA protein was detemined with phases obtained by isomorphous replacement and by

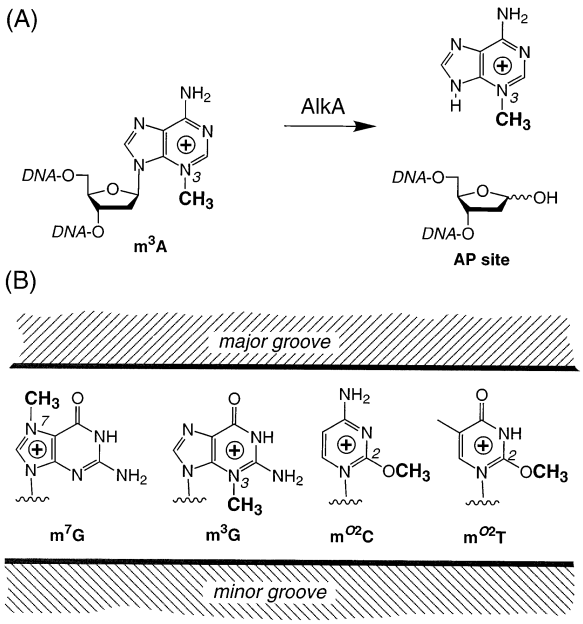



Figure 1. Repair of Aberrantly Methylated DNA by the E. coli AlkA Protein

(A) AlkA catalyzes hydrolysis of the glycosidic bond linking 3-methyladenine ($m^3A$) to the DNA backbone. The resulting abasic (AP) site is then repaired through excision/replacement DNA synthesis.

(B) Additional substrates that are processed efficiently by AlkA. AlkA substrates differ in their steric characteristics, arrangement of hydrogen-bonding functionality, and groove location of the aberrant methyl group (bold). A common feature of the substrates is their positive charge (Lindahl, 1982), which makes their $\pi$-surface electron deficient.

multiwavelength anomalous diffraction from crystals of selenomethionine-substituted AlkA protein (Table 1). The model of the two AlkA protomers (282 residues each) in the asymmetric unit has been refined to a crystallographic R-factor of 0.20 at 1.8 Å resolution ($R_{free}$ = 0.27) with tight restraints on protein geometry. Representative $2F_o$-$F_c$ difference electron density calculated from the refined structure is shown in Figure 2.

AlkA is a compact globular protein consisting of three roughly equal-sized domains with overall dimensions of approximately 50 Å by 45 Å by 25 Å. The amino-terminal 88 residues of the protein form a mixed $\alpha$-$\beta$ structure with a five-stranded antiparallel $\beta$ sheet flanked by two $\alpha$ helices (Figures 3A–3C). The shape and topology of domain 1 are very similar to those of the conserved tandem repeat of the TATA-binding protein (Nikolov et al., 1992; Y. Kim et al., 1993; J. Kim et al., 1993). This similarity in domain structure probably is a consequence of a preferred $\beta$-sheet/$\alpha$-helix packing arrangment rather than any common function of this fold in AlkA and TATA-binding protein. Helix $\alpha$A of domain 1 contacts the BC-loop and strand $\beta$D through hydrophobic interactions. Helix $\alpha$B lies across the curved surface of the $\beta$ sheet where it is anchored by hydrophobic interactions with the underlying strands. The exposed surface of helix B is decorated with charged and polar residues that face a surface depression bordered by $\alpha$B, $\alpha$D, and $\alpha$F (Figure 3A). The peptide segment connecting domains 1 and 2 of AlkA consists of the $\alpha$C helix (residues 89–99), which packs against domain 3, followed by an extended strand that is mostly buried in the protein interior (residues 100–112; Figures 3A and 3C).

Domain 2 is a globular bundle of seven $\alpha$ helices ($\alpha$D through $\alpha$J; residues 113–230) with an extensive hydrophobic core (Figures 3A–3C). Several prominent loops connect the helices of domain 2. A pendulous loop joining helices $\alpha$E and $\alpha$F inserts between the $\beta$-CD loop and strand $\beta$B of domain 1 (Figure 3B). This EF-loop has a left-handed twist that is stabilized by interstrand hydrogen bonds at the crossover point, and the tip of the EF-loop protrudes almost at right angles to the $\beta$-sheet surface, exposing three acidic residues to solvent (Asp-147, Asp-148, Glu-151). A second compound loop connecting $\alpha$H and $\alpha$I of domain 2 packs against the surfaces of $\alpha$D and $\alpha$J through nonpolar interactions (Figure 3A). Helices $\alpha$I and $\alpha$J and the intervening type II $\beta$ turn form a conserved motif termed the helix-hairpin-helix (HhH), which is the binding site for thymine glycol in crystals of Endo III (Thayer et al., 1995). The residues comprising the hairpin of the AlkA HhH, –Phe-212–Pro-213–Gly-214–Ile-215–, are held in place by the insertion of Ile-215 into a crevice between $\alpha$D and $\alpha$H. In addition, Phe-212 stacks against His-184 of $\alpha$H in an amino/aromatic interaction (Burley and Petsko, 1986; Mitchell et al., 1994). $\alpha$J of the HhH contributes the aromatic residues Trp-218 and Tyr-222 to the cleft between domains 2 and 3.

The C-terminal domain of AlkA is a right-handed three-helix bundle of $\alpha$K through $\alpha$M (residues 231–282) with one additional helix contributed by the connecting segment between domains 1 and 2 ($\alpha$C, residues 89–99; Figures 3A–3C). The $\alpha$C helix packs against $\alpha$M primarily through hydrophobic interactions, completing the four-helix bundle of domain 3. A prominent cleft separates domains 2 and 3 of AlkA, and few contacts are made between domains 1 and 3. Domain 1 serves as a platform that accepts the other two domains of the protein.

Nonpolar residues occupy most of the buried surface between domains 1 and 2 (1868 Å$^2$ of buried surface), with a few notable exceptions. Two buried arginine–glutamic acid salt bridges join $\alpha$E of domain 2 with strands $\beta$B and $\beta$C of domain 1. An extensive network of hydrogen bonds between these residue pairs fully satisfies their bonding potential within the hydrophobic core of the protein. One of these arginine-glutamic acid pairs is partially exposed to solvent via a small pocket in the protein located above helix $\alpha$B in Figure 3A. A cluster of water molecules occupies this pocket in both molecules of the crystallographic asymmetric unit. Several other charged residues decorate the rim of this surface depression, including the exposed residues of helix $\alpha$B (domain 1) and residues of the EF-loop (domain 2) described above. The interface between domain 3 and domain 1 is less extensive than that involving domains 2 and 1, suggestive of a flexible connection between domain 3 and the rest of the protein. Although most of the solvent-inaccessible surface of domain 3 consists of hydrophobic residues, a partially buried histidine of $\alpha$M (His-270) bridges between a main chain oxygen of $\alpha$C and a neighboring main chain oxygen of $\alpha$M. This electrostatic interaction further stabilizes the packing of $\alpha$C with the rest of domain 3.

Table 1. Structure Determination and Refinement

X-ray Diffraction Data

| Data Set | Source[a] | λ (Å) | No. Refl. | Unique | Completeness (%) | $R_{sym}$[b] | Resol. (Å) |
|---|---|---|---|---|---|---|---|
| Native | 1 | 1.542 | 318470 | 45971 | 99.8 | 0.051 | 1.80 |
| PIP | 2 | 1.542 | 24757 | 11752 | 83 | 0.044 | 2.85 |
| Se0 | 2 | 1.542 | 22165 | 10963 | 78 | 0.025 | 2.70 |
| Se1 | 3 | 0.9871 | 95685 | 32246 | 89 | 0.053 | 2.00 |
| Se2 | 3 | 0.9792 | 109295 | 30542 | 90 | 0.068 | 2.00 |
| Se3 | 3 | 0.9789 | 109546 | 31133 | 92 | 0.076 | 2.00 |
| Se4 | 3 | 0.9649 | 112689 | 30257 | 89 | 0.060 | 2.00 |

Phasing Statistics

| | | | $R_{Cullis}$[e] | | | Phasing Power[d] | |
|---|---|---|---|---|---|---|---|
| Derivative | Resol. (Å) | No. Sites | Acentric | Centric | Anomalous | Acentric | Centric |
| PIP | 16.0–4.9 | 4 | 0.65 | 0.65 | 0.80 | 1.4 | 1.3 |
| Se0 | 16.0–2.7 | 16 | 0.53 | 0.55 | — | 1.8 | 1.2 |
| Se1 | 8.5–2.0 | 16 | 0.65 | 0.71 | — | 1.6 | 0.9 |
| Se2 | 8.5–2.0 | 16 | 0.82 | 0.86 | 0.60 | 0.5 | 0.3 |
| Se3 | 8.5–2.0 | 16 | 0.72 | 0.77 | 0.50 | 0.9 | 0.5 |
| Se4 | 8.5–2.0 | 16 | 0.67 | 0.68 | 0.60 | 1.1 | 0.7 |

Figure of Merit (acentric/centric) 0.53/0.68

Model Refinement

| | Resol. (Å) | $R_{cryst}$[e] | $R_{free}$[e] | rms Bonds | rms Angles | B Correl. |
|---|---|---|---|---|---|---|
| X-PLOR | 6.0–1.8 | 0.218 | 0.273 | | | |
| TNT | 30.0–1.8 | 0.202 | 0.287 | 0.014 | 1.682 | 3.629 |

[a] 1: Laboratory X-ray source, Marresearch image plate; 2: Laboratory X-ray source, Siemens multiwire area detector; 3: Bean line X-4A, Brookhaven National Lab, Fuji image plates.
[b] $R_{sym} = \Sigma \mid Ij - < I > \mid / \Sigma < I >$, where *Ij* is the intensity of reflection j and $< I >$ is the mean intensity of multiple measurements.
[c] $R_{cullis} = \Sigma \parallel Fph \pm Fp \mid - Fh, c \mid / \Sigma \mid Fph \pm Fp \mid$, where *Fh, c* is the calculated heavy atom structure factor.
[d] Phasing Power $= < Fh > / E$, where $< Fh >$ is the root-mean-square heavy atom structure factor and E is the residual lack of closure error.
[e] $R_{cryst} = \Sigma \parallel Fo \mid - \mid Fc \parallel / \Sigma \mid Fo \mid$, calculated with the working reflection set. $R_{free}$ is the same parameter calculated with the reserved reflections.

## The Size of the Cleft between Domains 2 and 3 of AlkA is Adjustible

In heavy atom derivatives of AlkA crystals we observed a significant reorientation of domain 3 with respect to the rest of the protein (see Experimental Procedures). A hinge-like motion in the segment connecting αJ and αK allows domain 3 to rotate away from domain 2, widening the cleft between the two. Both protomers in the asymmetric unit undergo the same conformational transition in response to metal binding, despite their different crystal packing environments. Widening of the cleft between domains 2 and 3 upon binding of other ligands, most notably substrate bases, may provide additional flexibility in accomodating a range of alkylation adducts.

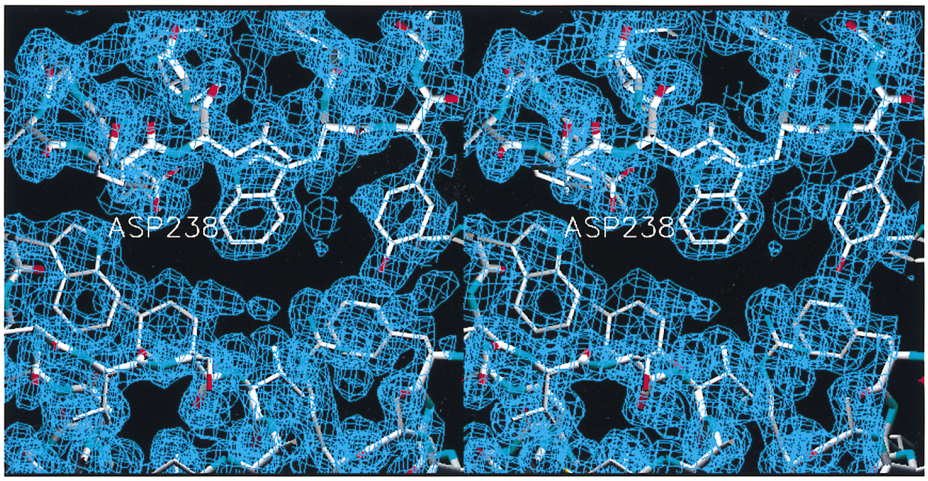



Figure 2. Stereo View of $2F_o$-$F_c$ Electron Density around the Aromatic Cleft Located between Domains 1 and 2 of the AlkA Protein

The map was calculated at 1.8 Å resolution using phases and amplitudes from the refined model. The catalytically essential Asp-238 residue is labeled. Figures were prepared with the program SETOR (Evans, 1990), except where otherwise noted.

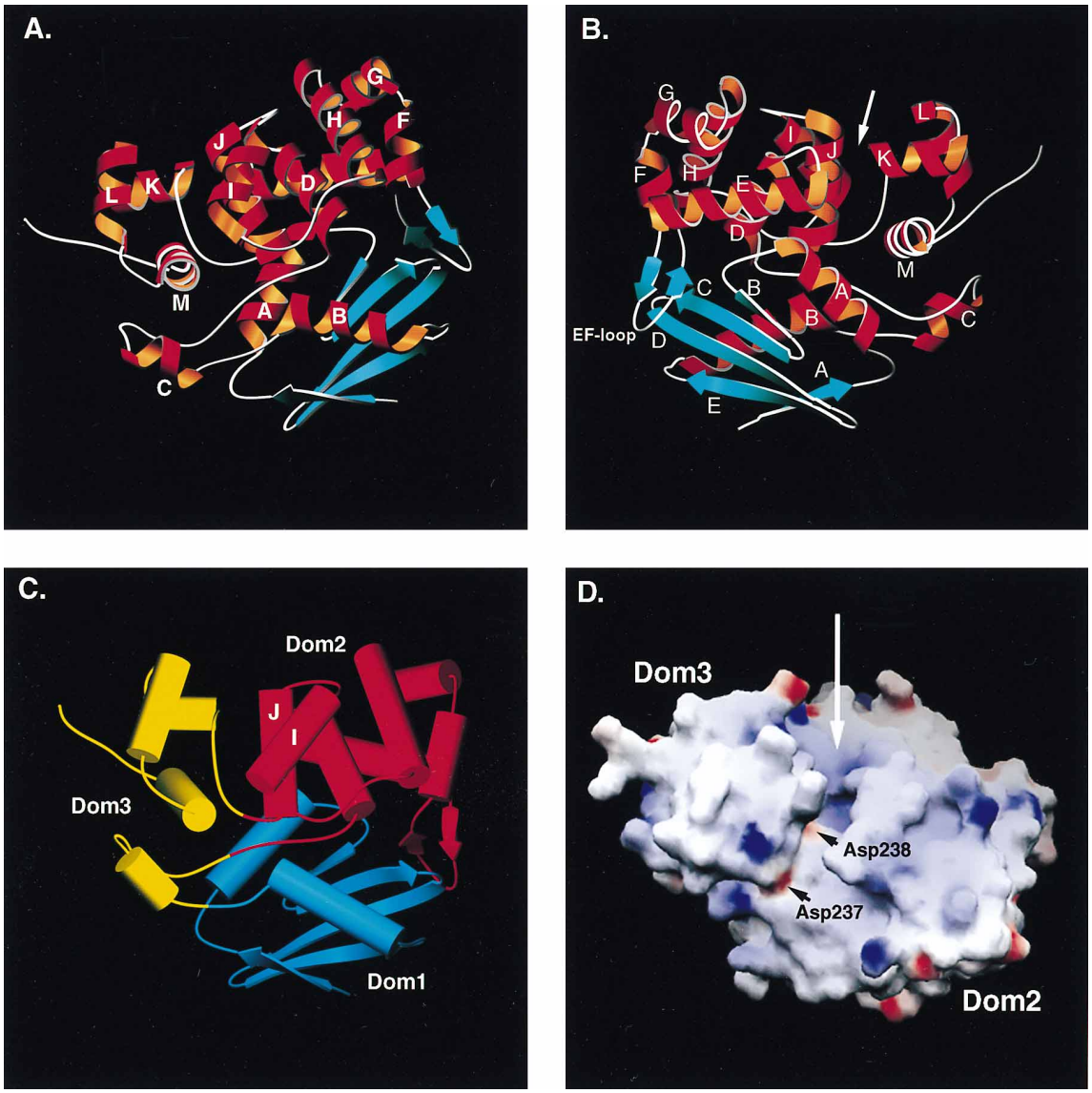



Figure 3. Overall Shape and Domain Structure of AlkA

(A and B) Course of the AlkA polypeptide chain, with elements of secondary structure assigned and colored accordingly (blue = β sheet, red/orange = α helix, white = nonrepetitive elements). The arrow in (A) shows the location of the proposed enzyme active site. The view in (B) is related to that in (A) by rotation of ~180°.

(C) The AlkA protein consists of three domains: an N-terminal mixed α-β structure (Domain 1, blue); a central seven-helix bundle (Domain 2, red; αD through αJ), and a C-terminal domain of four α helices (Domain 3, yellow; αC and αK through αM). A paucity of intersubunit contacts allows some movement of domain 3 with respect to the rest of the protein. The conserved helix-hairpin-helix motif, consisting of helices αI and αJ and the intervening β turn, is located on one side of this interdomain cleft.

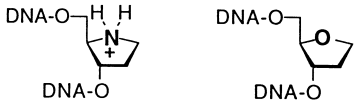
(D) The solvent-accessible surface of AlkA, colored according to electrostatic potential (blue, positively charged; red, negatively charged), reveals a cleft at the junction of domains 2 and 3, which is unusually rich in aromatic residues. Jutting into the cleft is the catalytically essential residue Asp-238. The neighboring Asp residue at position 237, which lies at the periphery of the aromatic cleft, is not essential for glycosylase activity. A number of lysines and arginines (blue), which could potentially interact with DNA backbone phosphates, decorate the protein surface around the aromatic cleft. This figure was created using the program GRASP (Nicholls et al., 1991).

Consistent with this notion, we have observed that equilibration of AlkA crystals with several methylated purine bases or nucleosides either cracks the crystals or induces significant changes in the unit cell dimensions.

### Catalytic Activity and DNA-Binding Properties of AlkA

Enzymes that catalyze glycosyl transfer reactions deliver an activated nucleophile to the anomeric carbon of the substrate (Davies and Henrissat, 1995). Such reactions typically take place in a cavity on the enzyme, so as to desolvate the reacting partners and align them stereochemically. The surface of AlkA contains two prominent cavities, a hydrophilic pocket containing the R37/E115 salt bridge (described above), and a hydrophobic cleft that lies at the interface of domains 2 and 3 (Figure 3D). Remarkably, almost 800 Å$^2$ of nonpolar surface area within the cleft is left exposed to solvent.

Table 2. Catalytic Activity and DNA-Binding Properties




| | Methylpurine Release (fmol) | | DNA-Binding Affinity (pM) | |
|---|---|---|---|---|
| | 8 min | 30 min | Pyrrolidine-DNA | Tetrahydrofuran-DNA |
| AlkA | 522 | 946 | 16 ± 3 | 45000 ± 6000 |
| D237N | 159 | 298 | 23 ± 9 | 12500 ± 8000 |
| D238N | 0 | 0 | 800 ± 500 | 1100 ± 400 |

The cleft also contains several charged residues, Arg-22, Asp-237, and Asp-238. Arg-22 and Asp-238 project into the hydrophobic cleft, whereas Asp-237 lies on the periphery (Figure 3D). Reasoning that a side-chain carboxylate might participate in catalysis, we mutated Asp-237 and Asp-238 individually to Asn and measured their glycosylase activity on DNA alkylated with [$^3$H]dimethylsulfate (Laval, 1977). Whereas the D237N mutant protein retains >30% of wild-type activity, substitution of Asp-238 with Asn (D238N) eliminates detectable catalytic activity (Table 2). To probe further the basis for catalytic inactivation by the D238N mutation, we measured binding to oligonucleotides containing modified abasic sites. The pyrrolidine AP-site is a potent mechanism-based inhibitor of AlkA (Schärer et al., 1995), which binds the wild-type protein with an affinity of 15 pM (Table 2). The D237N mutant of AlkA binds the pyrrolidine-AP site with similar affinity as wild-type AlkA. In contrast, the affinity of D238N AlkA for the pyrrolidine-AP site is reduced more than 20-fold. This decrease in affinity does not reflect a generalized loss of function, because the D238N mutant protein binds the uncharged tetrahydrofuran (THF)-AP analog with significantly increased affinity (Table 2). The pyrrolidine-AP site is believed to mimic the transition state for enzyme-catalyzed glycosidic bond cleavage (Schärer et al., 1995), and thus represents a specific probe of the enzyme active site. Indeed, the results are consistent with a direct electrostatic interaction between the positively charged pyrrolidine-AP analog and Asp-238. The enhanced affinity of the THF-site may result from the gain of a favorable interaction between the THF moiety and Asn-238. The significant alteration in binding specificity for AP analogs brought about by mutating Asp-238 to Asn provides strong evidence that this residue is located in the enzyme active site, most likely in close proximity to the sugar of the substrate.

It is likely that residues in the AlkA hydrophobic cleft nearby Asp-238 form a binding pocket for the methylated base (Figure 4). This model assumes that the substrate nucleoside is rotated out of the DNA helix and inserted into the active site cleft of the enzyme (see Figure 6), as has been observed with DNA methyltransferases (Klimasauskas et al., 1994; Reinisch et al., 1995) and postulated for uracil DNA glycosylases (Mol et al., 1995; Savva et al., 1995). The substrates for AlkA differ markedly in their steric properties and configuration of hydrogen-bonding functionality (see Figure 1); hence, it seems likely that the enzyme recognizes methylated bases through some alternative mechanism. The unifying feature of the alkylated bases processed efficiently

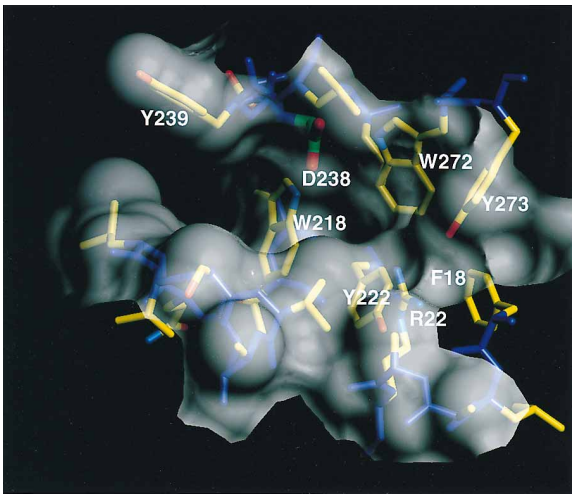



Figure 4. Detail of the Proposed Active Site of the AlkA Protein

The cleft, viewed along the direction of the arrow in Figure 3D, is rich in electron-donating aromatic side chains, which are well suited to recognize electron-deficient methylated bases through π–donor/acceptor interactions. The catalytically essential Asp-238 (green) lies at the bottom of the cleft, where it is poised to participate in the reaction chemistry, and to interact with mechanism-based oligonucleotide inhibitors.

by AlkA is their positive charge (Lindahl, 1982), which is distributed throughout the entire π-system of the base, making it electron-deficient. Such electron-deficient aromatic rings are known to form strong π–donor/acceptor (π–DA) interactions with electron-rich aromatic rings (Ishida et al., 1988). With this in mind, it is striking that the active site cleft of AlkA is lined with electron-rich aromatic rings contributed by the side-chains of Phe-18, Trp-218, Tyr-222, Trp-272, and Tyr-273 (Figure 4). Thus we propose that AlkA recognizes aberrently methylated DNA through π–DA interactions. This sterically insensitive mode of molecular recognition, together with flexion of the domain 2/3 interface, presumably allows AlkA to recognize a diverse group of modified bases.

## A Conserved Protein Fold for Base Excision-Repair

Genes encoding alkylpurine DNA glycosylases have cloned from a variety of organisms, including Bacillus subtilis (Morohoshi et al., 1993), Saccharomyces cerevisiae (Berdal et al., 1990; Chen et al., 1990), Schizosaccharomyces pombe (Memisoglu and Samson, 1996), Arabidopsis thaliana (Santerre and Britt, 1994), rat (O'Connor and Laval, 1990), mouse (Engelward et al., 1993) and humans (Chakravarti et al., 1991; O'Connor and Laval, 1991; Samson et al., 1991). Furthermore, the human and yeast enzymes complement E. coli strains deleted for the *alkA* and *tag* genes. Although the size of E. coli AlkA is similar to that of eukaryotic alkylpurine DNA glycosylases and these enzymes operate on many of the same substrates, their sequences bear little similarity. The sequence of the B. subtilis enzyme can be aligned reasonably well with that of AlkA, and the yeast MAG glycosylase has regions of limited homology; however, no sequence similarity is apparent between AlkA and the plant or mammalian enzymes.

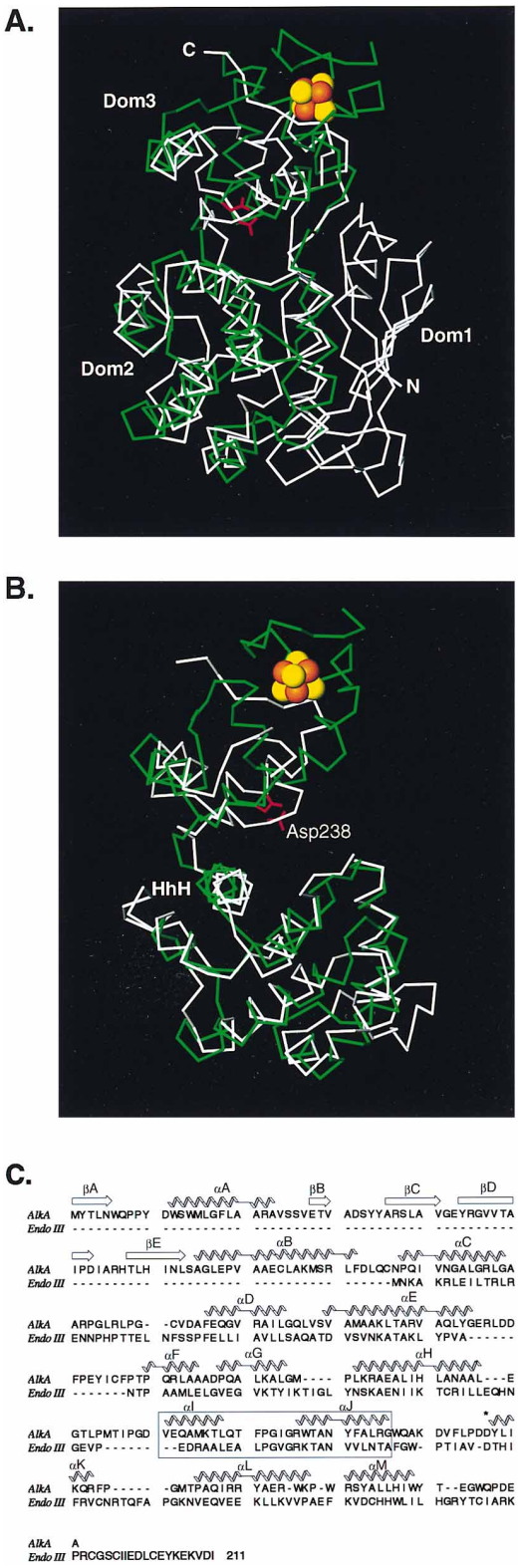



Figure 5. Structural Similarity of AlkA and E. coli Endonuclease III

(A and B) Cα trace of the AlkA protein (white) superimposed on that of the DNA glycosylase/lyase Endo III (green) (Thayer et al., 1995). (A) Much of domains 2 and 3 of AlkA correspond well with the two domains of Endo III, despite the nearly complete lack of amino-acid sequence similarity. Domain 1 of AlkA, which is structurally homologous to one conserved repeat of the TATA-binding protein, has no counterpart in Endo III. On the other hand, domain 3 of AlkA lacks an internal loop and C-terminal extension that in Endo III forms the protein scaffold for the [4Fe-4S] cluster (Fe, orange sphere; S, yellow sphere). (B) Cα superposition of the two homologous domains of AlkA, showing the structural homology of the conserved helix-hairpin-helix motifs of both proteins. The catalytically essential Asp residue of AlkA (Asp-238) aligns well with the corresponding catalytic Asp residue of Endo III (Asp-138). (C) Amino-acid sequence alignment based on the best superposition of Ca atoms in the two crystal structures, as determined by the distance geometry algorithm implemented in the program DALI (Holm and Sander, 1993). The secondary structure of AlkA is denoted above the sequence alignment. The highly conserved helix-hairpin-helix motif (HhH; boxed sequence in [C]) borders one side of the aromatic cleft of AlkA (refer to [B]). The HhH motif is present in a variety of DNA glycosylases and other DNA-binding proteins (Thayer et al., 1995; Doherty et al., 1996; Nash et al., 1996). On the other side of the AlkA interdomain cleft is the conserved aspartic acid required for DNA glycosylase activity (Asp-238; indicated by an asterisk in [C]).

The protein data bank was searched for structurally related proteins by comparing the interresidue (Cα–Cα) distances of AlkA with those in proteins of known structure (the DALI server; Holm and Sander, 1993). As noted above, the structure of domain 1 is similar to that of the conserved tandem repeat in TBP (Nikolov et al., 1992). A more striking similarity was found for domains 2 and 3 of AlkA, which correspond in fold and topology to a large portion of E. coli endonuclease III (Figure 5). The structural homology is especially strong in the HhH motif, a conserved element that is also present in a number of DNA glycosylases and other DNA-binding proteins (Thayer et al., 1995; Doherty et al. 1996). Just as with AlkA, residues within the interdomain cleft of Endo III are essential for catalytic activity (Thayer et al., 1995); Lys-120, which corresponds to Trp-218 of AlkA (Figures 5C and 6C), projects from helix 2 of the HhH into the interdomain cleft. Asp-138 of Endo III corresponds to the catalytically indispensible Asp-238 of AlkA (Figure 5). Most residues in the interdomain cleft of Endo III are charged and polar, which may reflect Endo III's preference for polar substrates.

A feature of the Endo III structure that has gained widespread attention is its [4Fe-4S] cluster, which is coordinated to four Cys residues in the C-terminal domain. This metal cluster has no apparent role in the reaction chemistry of Endo III, but rather is believed to stablize the protein fold (Kuo et al., 1992; Thompson, 1993; Thayer et al., 1995). Remarkably, the peptide segments that ligate the metal in Endo III are absent from AlkA, yet the two proteins adopt otherwise similar overall folds (Figures 6A and 6B). Another point of difference is the N-terminal domain of AlkA, which has no counterpart in Endo III. This domain walls off one end of the AlkA aromatic cleft, to which it contributes Phe-18 (Figure 4). It is notable that the S. pombe 3-methyladenine DNA glycosylase also lacks a segment homologous to the N-terminal TBP-like domain of E. coli AlkA protein (Memisoglu and Samson, 1996).

## Nucleophilic Activation by DNA Glycosylases and Glycosylase/Lyases

Two classes of base-excision DNA repair proteins are known, monofunctional glycosylases such as AlkA and

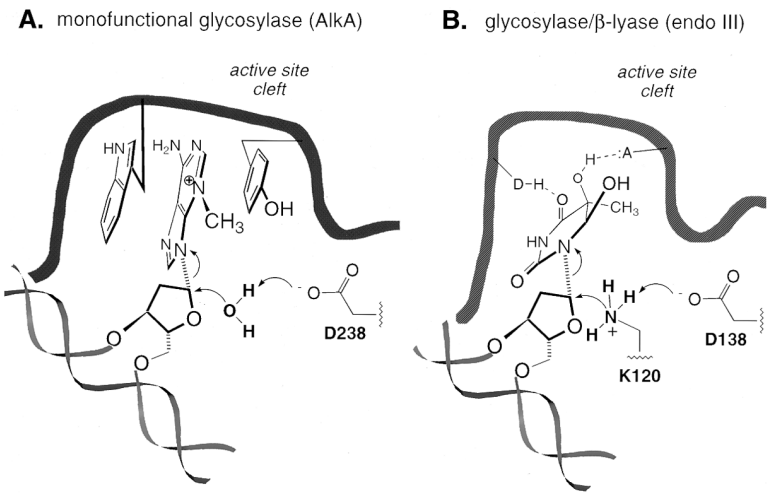



**Figure 6. Extrahelical Recognition and Excision of Aberrant DNA Bases by AlkA and Endo III**

The common location of a catalytically essential, conserved aspartic acid residue in the active site clefts of AlkA (A, Asp-238) and Endo III (B, Asp-138) suggests the two proteins may use a common mode of nucleophilic activation. In the case of AlkA, a monofunctional glycosylase, Asp-238 is envisioned to deprotonate water, activating it for attack on the glycosidic bond of the substrate. In the case of Endo III, we propose Asp-138 deprotonates the $\epsilon$-$NH_3^+$ group of Lys-120, which attacks the glycosidic bond to form a covalent enzyme-substrate intermediate (Dodson et al., 1994). The active site cleft of AlkA is rich in electron-donating aromatic residues (here notionally illustrated in [A] as a Trp and a Tyr residue), which are ideally suited to interact with electron-deficient bases (refer also to Figure 1B) through $\pi$-DA interactions. On the other hand, the corresponding cleft in Endo III contains mostly hydrophilic residues, which may interact with substrate bases such as thymine glycol (shown in [B]) through either direct or water-mediated hydrogen-bonding.

glycosylase/lyases such as Endo III. Although both classes release damaged bases from DNA, they generate different products. Monofunctional glycosylases produce an AP site (see Figure 1), whereas glycosylase/lyases further degrade the target sugar moiety by $\beta$ elimination of the 3′-phosphate. These seemingly profound differences in reaction pathway have led to the widespread belief that glycosylases and glycosylase/lyases may use fundamentally different catalytic mechanisms.

Evidence suggests that Endo III initiates base excision by attack of the Lys-120 $\epsilon$-$NH_2$ group on the substrate anomeric carbon (Figure 6) (Sun et al., 1995; Thayer et al., 1995). The covalent enzyme-DNA intermediate thus formed undergoes several subsequent transformations, ultimately resulting in hydrolytic degradation of deoxyribose and regeneration of the free enzyme (Dodson et al., 1994). In the active site cleft of Endo III, Asp-138 is well-positioned to deprotonate Lys-120, an obligate step for activation of the nucleophile. The reaction pathway of monofunctional glycosylases is less understood, but the failure to detect covalent enzyme-DNA intermediates suggests that water is the attacking nucleophile. The conserved positioning of a key catalytic Asp residue in the active site clefts of AlkA and Endo III suggests this residue may serve the same function in both classes of glycosylase. We therefore propose that Asp-238 of AlkA deprotonates water, thereby activating it for nucleophilic attack (Figure 6). These considerations point to a fundamental similarity in the reaction chemistry of glycosylases and glycosylase/lyases.

## Experimental Procedures

### AlkA Crystallization and X-ray Data Collection

The crystal structure of AlkA protein was determined by a combination of multiple isomorphous replacement (MIR) and multiple wavelength anomalous diffraction (MAD) methods. Single crystals of AlkA diffracting to Bragg spacings beyond 1.8 Å were obtained by vapor diffusion using protein purified from an overexpressing strain of E. coli. Pure AlkA protein (8–10 mg/ml) in 0.1 M NaCl, 0.02 M Tris–Cl (pH 7.5), 3 mM dithiothreitol, 0.1 mM EDTA was mixed with an equal volume of a well solution containing 0.015 M $KPO_4$ (pH 5.9) and 6%–10% (w/v) polyethylene glycol 8000. The crystals grow in hanging drops over the course of 2 to 7 days and diffraction quality is critically dependent upon enzyme purity and the precise protein concentration used in crystallization experiments. Crystals of AlkA protein belong to the monoclinic space group $P2_1$ (a = 58.22 Å, b = 75.96 Å, c = 61.03 Å; $\beta$ = 109.9°) with two AlkA protomers in the crystallographic asymmetric unit (Matthews coefficient = 2.18 $Å^3 da^{-1}$). The two molecules in the asymmetric unit pack in a head-to-tail arrangement that generates an approximate C-centered crystal symmetry that is evident from the very weak intensities for reflections of parity [h+k] = 2n+1.

### Heavy Atom Derivatization and Phasing

Crystals of AlkA protein react strongly with a wide variety of heavy metals. Metal exposure severely decreased the X-ray diffraction quality of the crystals or resulted in significant changes in crystal unit cell parameters, despite our efforts to limit metal substitution. Metal-induced changes in the average intensities of the [h+k] = 2n reflections in comparison to the [h+k] = 2n+1 reflections suggested that metal substitution affected the global packing arrangment of AlkA protomers in the crystals. We therefore produced crystals of selenomethionine-substituted AlkA protein using published methods (Van Duyne et al., 1993) and initial phases were obtained from multiwavelength anomalous diffraction (MAD) experiments using synchrotron X-radiation (beamline X-4A; National Synchrotron Light Source, Brookhaven National Laboratory). Twelve of the 16 seleniums in the crystal asymmetric unit were located by Patterson vector searches. Difference Fourier methods were used to locate the remaining 4 selenium atoms within the asymmetric unit. Additional single-wavelength diffraction data (Cu $K_\alpha$, $\lambda$ = 1.54 Å) were collected from the selenomethionine crystal used for the MAD experiments (Se0 data). Inclusion of this latter data set during heavy atom parameter refinement markedly improved the quality of the resulting phases, as evidenced by a decrease in the Cullis R-factor for the MAD data and improved electron density. Heavy atom parameter refinement for the five selenium data sets and an isomorphous native data set was performed by treating the MAD data as a special case of isomorphous replacement as previously described (Ramakrishnan et al., 1993; CCP4, 1979). Initial phases were improved by real space density modification (Zhang and Main, 1990; Cowtan and Main, 1993), and the resulting electron density calculated low resolution showed a clear solvent boundary and features of protein secondary structure. However, the initial density was disconnected and we were unable to trace the polypeptide chain. A search for additional heavy atom derivatives yielded a di-$\mu$-iodobis(ethylenediamine)diplatinum(II)nitrate derivative (PIP) that contributed enough low resolution phase information to trace most of the polypeptide chain. Phase modification by 2-fold averaging, solvent flattening,

and histogram matching yielded a significant improvement in the protein main-chain density calculated at low resolution. Phase modification by application of the Sayre relationship (Main, 1990) to the high resolution MAD phases markedly improved the electron density corresponding to protein side chains, although these maps lacked the continuity of those calculated with the combined MIRAS/MAD phase information. A model of the two AlkA protomers was built into the electron density using the graphics program "O" (Jones, 1992).

### Model Refinement

Partial structures of the two AlkA protomers were refined using X-PLOR (Brünger, 1992a) or TNT (Tronrud, 1995), initially with the application of noncrystallographic restraints to both main-chain and side-chain atoms. In the later stages, model refinement by simulated annealing was used and NCS restraints were applied only to main-chain atoms. The success of model refinement was evaluated at each stage from changes in the free R-factor (Brünger, 1992b) and the validity of the refined structure was confirmed by simulated annealing-omit procedures (Brünger, 1992a) and by profile analysis of the final refined structure (Lüthy et al., 1992). Proper refinement of the protein model depended upon including most of the systematically weak [h+k] = 2n+1 reflections in the calculations. Refinement using only reflections with $F/\sigma \geq 3.0$ (which excluded many of the [h+k] = 2n+1 reflections) resulted in a higher free R-factor, particularly for reflections in the high resolution bins. The program TNT was used in the final stages of model refinement, with the inclusion of all reflections from 30 Å to 1.8 Å resolution. R-factors listed in Table 1 are for models of 2 AlkA protomers (282 residues each) with either 300 waters (X-PLOR) or 359 waters (TNT) and they are calculated for all observed reflections. The root-mean-square deviations in bond lengths and bond angles are as reported by TNT, and the atomic B-factor correlation is that calculated using X-PLOR.

### Glycosylase Activity Assay

Base-excision activity was assayed by mixing 800 fmol of pure AlkA protein with ~2 pmol of calf thymus DNA that had been treated with [$^3$H]dimethylsulfate in HEPES buffer (pH 7.8) containing 1 mM EDTA and 1 mM dithiothreitol. The reaction was stopped with 0.25 volumes of 1M NaCl solution containing 1 mg ml$^{-1}$ bovine serum albumin and 1 mg ml$^{-1}$ salmon sperm DNA, and the DNA was precipitated with two volumes of ethanol. The amount of radiolabeled base remaining in the ethanol supernatant was measured by scintillation counting. The apparent DNA-binding affinities were measured by electrophoretic gel mobility shift assay for DNAs containing either the pyrrolidine abasic site or the tetrahydrofuran abasic site as described (Schärer et al., 1995).

### Acknowledgments

Correspondence should be addressed to G. L. V. or to T. E. E. We thank Craig Ogata and Wayne Hendrickson for assistance with MAD data collection at beamline X-4A, National Synchrotron Light Source (Upton, NY); Alan Friedman for providing local scaling and MAD data analysis programs; and members of the Ellenberger, Verdine, Hogle, and Samson research groups for stimulating discussions. This work was supported by grants from the National Institutes of Health (T. E. E./G. L. V.) and the Lucille P. Markey Charitable Trust (T. E. E.). G. L. V. is a Presidential Young Investigator and T. E. E. is a Lucille P. Markey Scholar in Biomedical Sciences.

Received June 11, 1996; revised July 5, 1996.

### References

Berdal, K.G., Bjørås, M., Bjelland, S., and Seeberg, E. (1990). Cloning and expression in *Escherichia coli* of a gene for an alkylbase DNA glycosylase from *Saccharomyces cerevisiae*; A homologue to the bacterial *alkA* gene. EMBO J. *9*, 4563–4568.

Brünger, A.T. (1992a). X-PLOR Version 3.1. A System for X-ray Crystallography and NMR (New Haven, Connecticut: Yale University Press).

Brünger, A.T. (1992b). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature *355*, 472–475.

Burley, S.K., and Petsko, G.A. (1986). Amino-aromatic interactions in proteins. FEBS Letters *203*, 139–143.

CCP4. (1979). The SERC (UK) Collaborative Computing Project No. 4. A Suite of Programs for Protein Crystallography. (Warrington, UK: Daresbury Laboratory).

Chakravarti, D., Ibeanu, G.C., Tano, K., and Mitra, S. (1991). Cloning and expression in *Escherichia coli* of a human cDNA encoding the DNA repair protein N-methylpurine-DNA glycosylase. J. Biol. Chem. *266*, 15710–15715.

Chen, J., Derfler, B., and Samson, L. (1990). *Saccharomyces cerevisiae* 3-methyladenine DNA glycosylase has homology to the AlkA glycosylase of E. coli and is induced in response to DNA alkylation damage. EMBO J. *9*, 4569–4575.

Cowtan, K.D., and Main, P. (1993). Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints. Acta Crystallogr. *D49*, 148–157.

Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolyases. Structure *3*, 853–859.

Dodson, M.L., Michaels, M.L., and Lloyd, R.S. (1994). Unified catalytic mechanism for DNA glycosylases. J. Biol. Chem. *269*, 32709–32712.

Doherty, A.J., Serpell, L.C., and Ponting C.P. (1996). The Helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA. Nucleic Acids Res. *24*, 2488–2497.

Engelward, B.P., Boosalis, M.S., Chen, B.J., Deng, Z., Siciliano, M.J., and Samson, L.D. (1993). Cloning and characterization of a mouse 3-methyladenine/7-methylguanine/3-methylguanine DNA glycosylase cDNA whose gene maps to chromosome 11. Carcinogenesis *14*, 175–181.

Evans, S.V. (1990). SETOR: hardware lighted three-dimensional solid model representations of macromolecules. J. Mol. Graphics *11*, 134–138.

Friedberg, E.C., Walker, G.C., and Siede, W. (1995). DNA Repair and Mutagenesis (Washington, D.C.: Am. Soc. Microbiol.).

Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J. Mol. Biol. *233*, 123–138.

Ishida, T., Doi, M., Ueda, H., Inoue, M., and Scheldrick, G.M. (1988). Specific ring stacking interaction on the tryptophan-7-methylguanine system: comparative crystallographic studies of indole derivatives-7-methylguanine base, nucleoside, and nucleotide complexes. J. Am. Chem. Soc. *110*, 2286–2294.

Jones, T.A., and Kjeldgaard, M. (1992). O—The Manual (Uppsala, Sweden: http://kaktus.kemi.aau.dk).

Klimasauskas, S., Kumar, S., Roberts, R.J., and Cheng, X. (1994). HhaI methyltransferase flips its target base out of the DNA helix. Cell *76*, 357–369.

Kim, J., Nikolov, D.B., and Burley, S.K. (1993). Crystal structure of TBP recognizing the minor groove of a TATA element. Nature *365*, 520–527.

Kim, Y., Geiger, J.H., Hahn, S., and Sigler, P.B. (1993). Crystal structure of a yeast TBP/TATA-box complex. Nature *365*, 512–520.

Kuo, C.-F., McRee, D.E., Fisher, C.L., O'Handley, S.F., Cunningham, R.P., and Tainer, J.A. (1992). Atomic structure of the DNA repair [4Fe-4S] enzyme endonuclease III. Science *258*, 434–440.

Laval, J. (1977). Two enzymes are required for strand incision in repair of alkylated DNA. Nature *269*, 829–832.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. Nature *362*, 709–715.

Lindahl, T. (1982). DNA repair enzymes. Annu. Rev. Biochem. *51*, 61–87.

Lüthy, R., Bowie, J.U., and Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. Nature *356*, 83–85

Main, P. (1990). The use of sayre's equation with constraints for the direct determination of phases. Acta Crystallogr. *A46*, 372–377.

Mattes, W.B., Lee, C.-S., Laval, J., and O'Connor, T.R. (1996). Excision of DNA adducts of nitrogen mustards by bacterial and mammalian 3-methyladenine-DNA glycosylases. Carcinogenesis *17*, 643–648.

Memisoglu, A., and Samson, L. (1996). Cloning and characterization of a cDNA encoding a 3-methyladenine DNA glycosylase from the fission yeast *Schizosaccharomyces pombe*. Gene, in press.

Mol, C.D., Arvai, A.S., Slupphaug, G., Kavil, B., Alseth, I., Krokan, H.E., and Tainer, J.A. (1995). Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. Cell *80*, 869–878.

Morohoshi, F., Hayashi, K., and Munkata, N. (1993). *Bacillus subtilis alkA* gene encoding inducible 3-methyladenine DNA glycosylase is adjacent to the *ada* operon. J. Bacteriol. *175*, 6010–6017.

Mitchell, J.B.O., Nandi, C.L., McDonald, I.K., and Thornton, J.M. (1994). Amino/aromatic interactions in proteins: is the evidence stacked against hydrogen bonding? J. Mol. Biol. *239*, 315–331.

Nash, H.M., Bruner, S.E., Schärer, O.D., Kawate, T., Addona, T.A., Spooner, E., Lane, W.S., and Verdine, G.L. (1996). Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily. Curr. Biol., in press.

Nicholls, A., Sharp, B.A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins *11*, 281–296.

Nikolov, D.B, Hu, S.-H., Lin, J., Gasch, A., Hoffman, A., Horikoshi, M., Chua, N.-H., Roeder, R.G., and Burley, S.K. (1992). Crystal structure of TFIID TATA-box binding protein. Nature *360*, 40–46.

O'Connor, T.R., and Laval, J. (1990). Isolation and structure of a cDNA expressing a mammalian 3-methyladenine-DNA glycosylase. EMBO J. *9*, 3337–3342.

O'Connor, T., and Laval, J. (1991). Human cDNA expressing a functional DNA glycosylase excising 3-methyladenine and 7-methylguanine. Biochem. Biophys., Res. Comm. *176*, 1170–1177.

Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L., and Sweet, R.M. (1993). Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. Nature *362*, 219–223.

Reinisch, K., 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.

Rydberg, B., and Lindahl, T. (1982). Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. EMBO J. *1*, 211–216.

Sakumi, K., and Sekiguchi, M. (1990). Structures and functions of DNA glycosylases. Mutation Res. *236*, 161–172.

Samson, L., Derfler, B., Boosalis, M., and Call, K. (1991). Cloning and characterization of a 3-methyladenine DNA glycosylase cDNA from human cells whose gene maps to chromosome 16. Proc. Natl. Acad. Sci. USA *88*, 9127–9131.

Santerre, A., and Britt, A.B. (1994). Cloning of a 3-methyladenine-DNA glycosylase from *Arabidopsis thaliana.* Proc. Natl. Acad. Sci. USA *91*, 2240–2244.

Savva, R., McAuley-Hecht, K., Brown, T., and Pearl, L. (1995). The structural basis of specific base-excision repair by uracil-DNA glycosylase. Nature *373*, 487–493.

Schärer, O.D., Ortholand, J.-Y., Ganesan, A., Ezaz-Nikpay, K., and Verdine, G.L. (1995). Specific binding of the DNA repair enzyme AlkA to a pyrrolidine-based inhibitor. J. Am. Chem. Soc. *117*, 6623–6624.

Seeberg, E., Eide, L., and Bjφrås, M. (1995). The base excision repair pathway. Trends Biochem. Sci. *20*, 391–397.

Sun, B., Latham, K.A., Dodson, M.L., and Lloyd, R.S. (1995). Studies on the catalytic mechanism of five DNA glycosylases. J. Biol Chem. *270*, 19501–19508.

Thayer, M.M., Ahern, H., Xing, D., Cunningham, R.P., and Tainer, J.A. (1995). Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. EMBO J. *14*, 4108–4120.

Thomas, L., Yang, C.-H., and Goldthwait, D.A. (1982). Two DNA glycosylases in *Escherichia coli* which release primarily 3-methyladenine. Biochemistry *21*, 1162–1169.

Thompson, A.J. (1993). Crosslinked by a cluster. Current Biol. *3*, 173–174.

Tronrud, D.E. (1995). User's guide, the TNT refinement package release 5E (Oregon State Board of Higher Education).

Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L., and Clardy, J. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. J. Mol. Biol. *229*, 104–125.

Vassylyev, D.G., Kashiwagi, T., Mikami, Y., Ariyoshi, M., Iwai, S., Ohtsuka, E., and Morikawa, K. (1995). Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: structural basis for damaged DNA recognition. Cell *83*, 773–782.

Zhang, K., and Main, P. (1990). Histogram matching as a new density modification technique for phase refinement and extension of protein molecules. Acta Crystallogr. *A46*, 41–46.