How do DNA repair proteins locate damaged bases in the genome?

Gregory L Verdine and Steven D Bruner

The genome is susceptible to the attack of reactive species that chemically modify the bases of DNA. If genetic information is to be transmitted faithfully to successive generations, it is essential that the genome be repaired. All organisms express proteins specifically dedicated to this task. But how do these proteins find the aberrant bases amongst the enormous number of normal ones?

Address: Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA.

Correspondence: Gregory L Verdine

Chemistry & Biology May 1997, 4:329–334
http://biomednet.com/elecref/1074552100400329

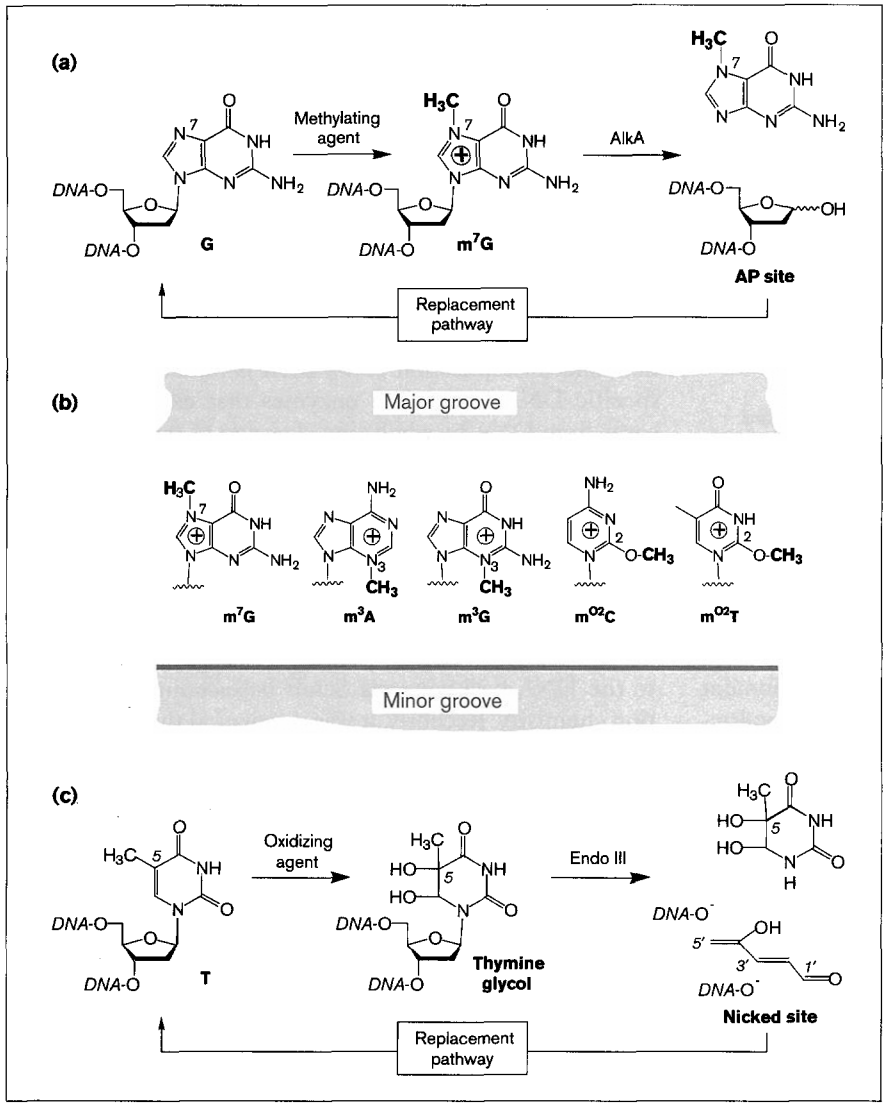
The genome is constantly under siege. Hardly a moment goes by in which genomic DNA does not suffer some irreversible change in its covalent structure due to attack, either by endogenous chemicals in the intracellular milieu or by exogenous agents that penetrate cells and irreversibly alter their contents [1]. Even water, the very molecule that endows DNA with its double-helical structure and enables it to interact with proteins, attacks DNA and in so doing alters the genetic code. Although most deleterious reactions that occur in the genome are hopelessly inefficient by the standards of preparative organic chemistry, the sheer size of chromosomes — each a single molecule of DNA that can comprise many billions of atoms — renders them unusually susceptible to even chance events. The heterocyclic bases of DNA are principal targets for modification by a wide variety of reactive species [1,2]. The molecular underpinnings of heredity are encoded within the chemical structure of these bases, so it stands to reason that certain covalent changes in base structure could affect their ability to be read by DNA and RNA polymerases, or worse yet, could affect their coding identity while they are being read. Indeed, known base modifications fall into both categories: those that are toxic by virtue of their ability to block replication and transcription, and those that are mutagenic because they miscode during replication. To counter the persistent threat posed by genotoxic lesions in DNA, all known free-living organisms (and even many viruses) express proteins that function solely to recognize and repair structural aberrations in their genomes [2]. Here we ponder the question of how DNA repair proteins, and in particular DNA glycosylases, locate their target lesions amidst the vast excess of normal genetic material.

The eradication of aberrant nucleobases from the genome is carried out by two systems, nucleotide excision repair (NER) and base excision repair (BER) [2]. Most spontaneously formed DNA lesions are repaired by the BER pathway. Thus, for example, BER is responsible for protecting the genome from aberrant bases that have arisen through UV cross-linking, oxidation, hydrolysis, alkylation, and misreplication [3–9]. BER is initiated by lesion-specific DNA glycosylase enzymes that excise damaged bases from DNA by catalyzing cleavage of their glycosidic bond. Two distinct classes of these enzymes are known. The first class, monofunctional DNA glycosylases, jettison the aberrant base via delivery of an activated water molecule to the glycosidic bond [10,11]. The second class, glycosylase/lyases, cleave the glycosidic bond through the attack of a enzymic amine nucleophile [9,12–15], and then further degrade the sugar moiety and sever its connections to the DNA backbone via Schiff base/conjugate elimination chemistry. Recently, it was discovered that a large proportion of the known DNA glycosylases from both families belong to a structural superfamily of BER proteins, members of which share a common core fold encompassing the enzyme active site [4].

The hallmark of BER superfamily proteins is an amino acid sequence motif comprising a helix–hairpin–helix (HhH) element [16] followed by a glycine- and/or proline-rich loop and an absolutely conserved, catalytically essential aspartate residue (HhH–G/PD motif) [4]. Interestingly, the HhH element alone has been suggested to occur in a wide variety of proteins that recognize unusual DNA structures, suggesting that perhaps there is a broader role for the HhH element in noncatalytic DNA-binding proteins [15,17]. The X-ray crystal structures of two prototypical members of the BER superfamily have been reported: endonuclease III (endo III), a glycosylase/lyase that processes oxidized pyrimidines in DNA [16], and AlkA, a monofunctional glycosylase that acts on aberrantly methylated base lesions [10,11] (Figure 1).

How do DNA glycosylases recognize their substrates? DNA glycosylases face the difficult problem of catalyzing a chemical reaction on a center — the sugar 1′-carbon of the substrate lesion — that is almost completely buried (and therefore inaccessible) in B-form DNA. This problem is akin to that confronted by DNA cytosine-5 methyltransferases (DCMtases), which must transfer a methyl group from *S*-adenosylmethionine to the C5-carbon of cytosine, a position that is occluded by the neighboring bases in DNA [18]. As revealed by X-ray co-crystallographic studies,

Figure 1

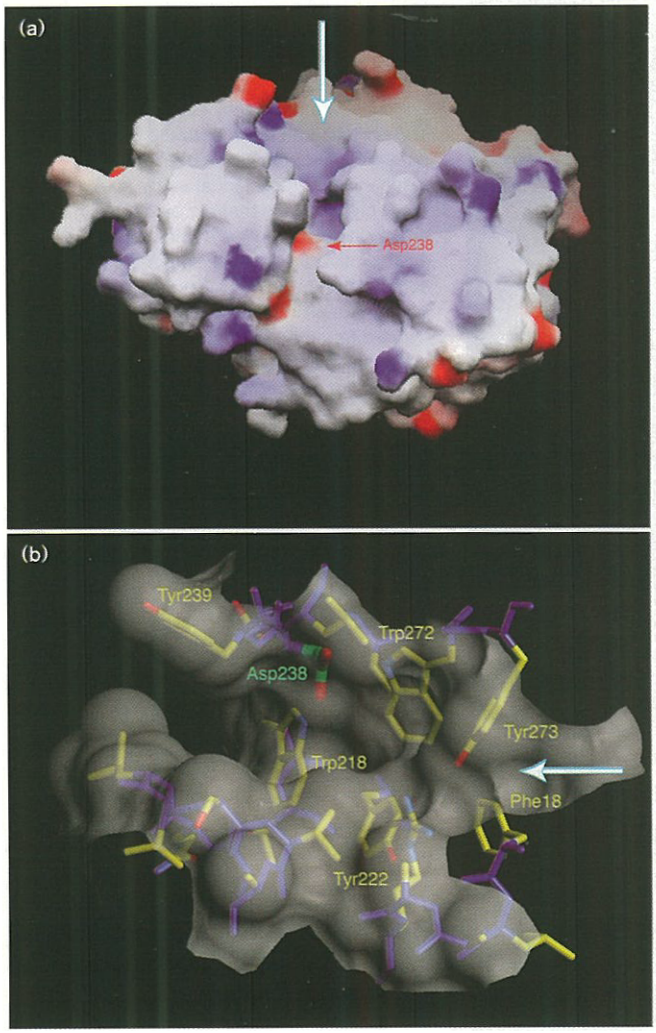


Generation and repair of aberrant bases in DNA. **(a)** A predominant site of DNA methylation is the N7-position of guanine. In bacteria, the adduct m^7G is processed by the enzyme AlkA, a monofunctional DNA glycosylase that catalyzes hydrolysis of the glycosidic bond. The original guanine residue is restored through a multienzyme system that removes the abasic site and resynthesizes the excised patch. **(b)** Adducts that are efficiently processed by AlkA. Note the lack of any common functionality or groove location of the methyl group amongst AlkA substrates. **(c)** Oxidation of thymine residues in DNA generates thymine glycol, a substrate for the DNA glycosylase/lyase endonuclease III (endo III). Endo III not only catalyzes cleavage of the glycosidic bond and jettisoning of the aberrant base, but also subsequent conjugate cleavage of the 5′- and 3′-phosphodiester linkage, so as to carve out the entire aberrant nucleoside unit from the DNA backbone. AlkA and endo III are members of a structural superfamily known as the base-excision DNA repair (BER) superfamily [4]. AlkA and endo III homologs are known in eukaryotes, including man.

DCMtases gain access to their substrate nucleoside by extruding it from the DNA helix; the resulting extrahelical base is then inserted into an active site pocket on the enzyme [19,20]. The discovery of extrahelical catalysis by DCMtases gave rise to speculation that base-specific DNA repair proteins would use a similar mechanism [18,19].

The first evidence in support of this hypothesis came from an examination of the co-crystal structure of (mutant) human uracil DNA glycosylase bound to its substrate, which revealed that the substrate uridine nucleoside is fully extrahelical and is bound deep within a concave pocket on the enzyme [21]. Although no co-crystal structure of a BER superfamily member bound to its substrate has yet been reported, there is every reason to believe that these proteins also employ an extrahelical mechanism for substrate recognition and processing. Indeed, the active site of AlkA is a prominent cleft lined by aromatic residues believed to recognize the substrate base through π-donor/acceptor (π-stacking) interactions (Figure 2; [10,11]). From the very bottom of this cleft juts the invariant, catalytically essential aspartate of the HhH-G/PD motif (Asp238), which is believed to activate the attacking nucleophile [10,11]. This active site cleft on the enzyme appears large enough to accept only a single nucleoside unit, and the substrate nucleoside must penetrate deeply into the cleft in order to position itself near Asp238. The existence of homology between endo III and AlkA suggests that the two enzymes possess a similar active site cleft, although the endo III active site is lined by fewer hydrophobic residues than that of AlkA [15,16]. In summary, a cornerstone of our model is the contention that most DNA glycosylases — certainly those that act on single-base lesions — bind their substrate nucleosides in an extrahelical active site.

Figure 2



The structure of AlkA (adapted from [10]; see also [11]). **(a)** Electrostatic surface (GRASP) representation of the AlkA surface, with negatively charged atoms in red, positively charged atoms in blue, and neutral atoms in white. Arrows indicate the active site cleft (white arrow) and the key catalytic aspartate residue, Asp238 (red arrow). **(b)** Solvent-accessible surface rendering of the AlkA active site, with residues presumed to be involved in base recognition indicated in yellow and the Asp238 indicated in green. The white arrow is drawn from roughly the same perspective as the white arrow in (a).

An interesting exception to the principle discussed above has been found in the case of phage T4 endonuclease V (endo V). Curiously, the co-crystal structure of endo V bound to DNA containing a thymine dimer reveals that although the thymine dimer substrate remains intrahelical, a deoxyadenosine nucleoside opposite the lesion is extrahelical [22]. The ability of endo V to avoid extrahelical catalysis may simply reflect the fact that thymine dimers disrupt duplex DNA structure over an unusually long distance (~10 Å), thus enabling the enzyme to invade the DNA helix, rather than vice versa.

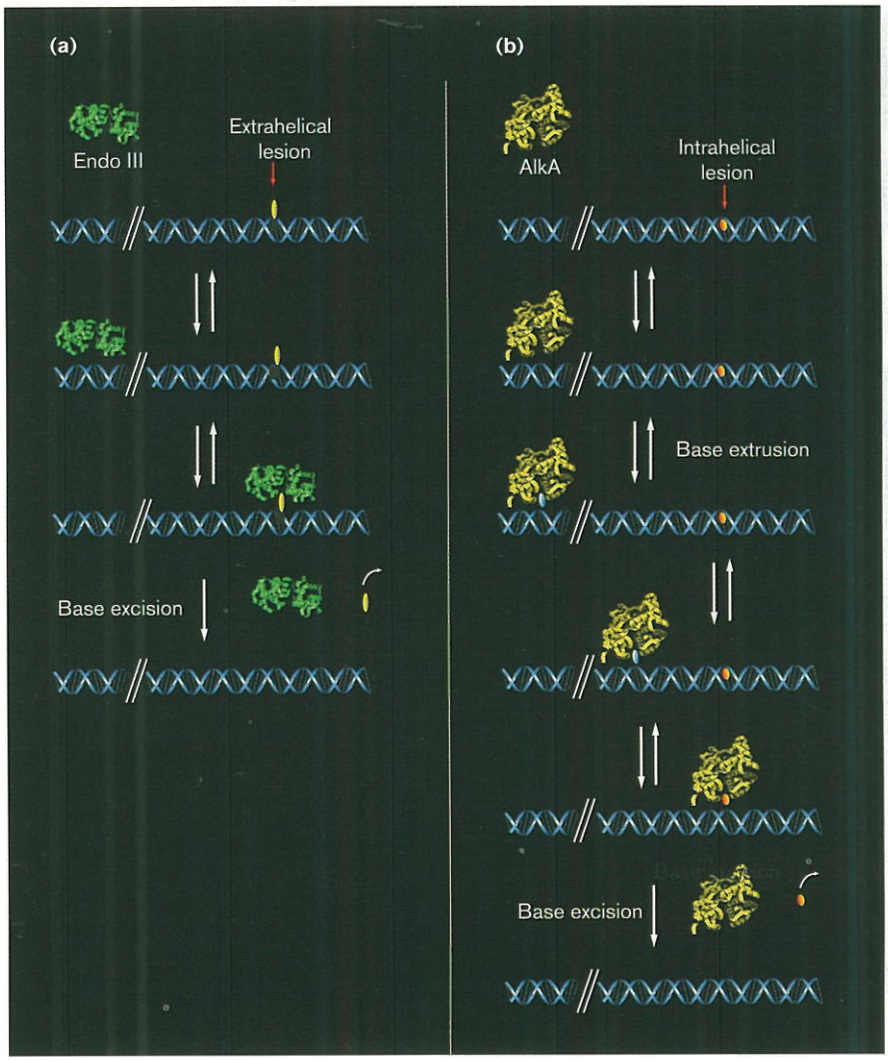
How do DNA glycosylases locate their target lesions in duplex DNA? Doing so represents a high-stakes version of finding a needle in a haystack; on the basis of the frequency at which spontaneously formed lesions arise in mammalian cells, one can surmise that DNA glycosylases must survey at least 10,000–100,000 nucleotides in order to locate a single lesion [2]. Furthermore, there is an especially high premium on the kinetic efficiency of repair, particularly in rapidly dividing cells, because the organism avoids deleterious biologic consequences only when repair precedes DNA replication.

Some modified bases, such as thymine glycol [23], substantially perturb the architecture and conformational dynamics of duplex DNA. It is not difficult to imagine how such lesions might cause profound dysfunction of DNA as a template, nor to envisage how the repair apparatus might recognize such egregiously non-native structures. Thymine glycol, like other helix-destabilizing base lesions, appears to interconvert readily between intrahelical and extrahelical orientations at room temperature [23]. Depending on the nature of the helical defect and the sequence context in which it occurs, the extrahelical conformation may even be thermodynamically preferred [24]. DNA glycosylases must ordinarily invest energy to induce duplex distortion, hence they might be expected to bind preferentially to substrates in which the cost of distortion is 'prepaid'. Indeed, most DNA glycosylases studied so far bind strongly to abasic site analogs [4,25–27] which, for the purposes of the present discussion, can be regarded as generic forms of damaged bases that destabilize duplex DNA structure. These considerations suggest a model for target location of DNA-destabilizing substrates by glycosylases such as endo III (Figure 3a). We imagine that these enzymes, which have basic DNA-recognition surfaces, track along the surface of DNA under the attractive influence of simple electrostatic forces [28]. While scanning the surface of DNA, the glycosylase encounters a locally destabilized site containing a base lesion that is perhaps already in an extrahelical orientation. The protein then locks its active site onto the lesion and the catalytic chemistry ensues.

Importantly, however, not all lesions that are efficiently processed by DNA glycosylases destabilize the double helix. One example of such a lesion is N7-methylguanine (m^7G), a toxic adduct generated through nonenzymic methylation of guanine in DNA. Biophysical studies have established that m^7G causes little if any disturbance of base-pair strength or duplex DNA structure [29,30], yet m^7G is efficiently recognized and processed by AlkA (Figure 1). Other prominent examples of nondestabilizing substrates include uracil, 8-oxoguanine and adenine (opposite A, C and 8-oxoguanine, respectively; [31]).

One possibility for target location in such cases would be that the enzyme scans DNA through electrostatic

Figure 3



(a) Proposed mechanism of target location by endo III and other glycosylases that recognize helix-destabilizing lesions. In this mechanism, the protein scans the surface of DNA until it encounters a transiently extrahelical lesion. The protein then locks onto the lesion and catalyzes base excision. **(b)** Proposed mechanism of target location by AlkA and other glycosylases that act on lesions that do not strongly destabilize duplex structure. We envision that the protein extrudes a base at some distance from the actual site of the lesion. The protein and extrahelical base then migrate together processively along the DNA helix until the aberrant base encounters the enzyme active site, whereupon catalysis ensues.

association until it experiences a chance encounter with an extrahelical lesion. Such encounters would be rare, as the rate of spontaneous extrusion for a nondestabilizing lesion is probably quite slow, and is certainly much slower than the rate of base-pair 'breathing' as detected by imino proton exchange (1–100 ms; [32]). Instead, we favor a scenario in which the enzyme plays a more active role in extruding the damaged base from the helix and presenting it to the active site. Potentially, the enzyme might recognize the non-native functionality of the lesion while it still lies in the DNA helix, and might then facilitate extrusion and subsequent processing of the lesion. Thus, for example, AlkA might recognize the presence of a non-native methyl group on the 7-position of guanine (m^7G; Figure 1), which would lie against the floor of the major groove. This would be no small feat, because the only recognition handles available to the enzyme in the major groove would be the rather sparse functionality that lies along the edge of the m^7G•C base pair (a methyl group and carbonyl oxygen on m^7G, and an aromatic amino-NH on C). More problematic is the fact that this mechanism would require that AlkA, a 31 kDa protein, possess specific interaction surfaces for methylated bases in the major (m^7G) and minor (m^3A, m^3G) grooves, and for chemically dissimilar adducts that occupy the same groove (m^3A and m^3G), all in addition to the common active site for processing extrahelical substrates. On the other hand, a single specificity pocket is capable of recognizing these substrates when bound in the aromatic active site cleft. We thus strongly disfavor the possibility that DNA glycosylases have distinct recognition surfaces for intrahelical and extrahelical target location, and instead propose that these enzymes select their substrates on the basis of chemical complementarity to a single extrahelical active site.

What is the most efficient way for a DNA glycosylase to sample all bases in DNA using an extrahelical active site?

Two limiting mechanisms, which we term nonprocessive and processive extrusion, are worthy of consideration. In the nonprocessive mechanism, the glycosylase randomly forces the extrusion of bases from the DNA helix, inserts them into the extrahelical active site, and either processes them (substrate bases) or allows them to reanneal (nonsubstrate bases). In the processive mechanism (Figure 3b), the enzyme also randomly targets a base for extrusion, but then the enzyme migrates, together with the extrahelical base, along the DNA helix, exchanging one extrahelical base for the next, until finally the enzyme happens upon a substrate lesion and catalysis ensues.

In the absence of experimental data supporting one extrusion mechanism over the other, is there good reason to favor either? We believe so. Enzymes that remodel DNA structure during the catalytic cycle must invest substantial energy in conformational reorganization of the substrate, such that the noncovalent chemistry can dominate the overall kinetics of the process [33]. This being the case, it would be exceedingly inefficient for a glycosylase to search for its substrates by the nonprocessive mechanism, in which the enzyme must pay the costly price of extrusion with each and every base sampling. Whereas *de novo* generation of an extrahelical base is energetically costly, migration of a pre-existing extrahelical base may proceed with little investment of additional energy because it represents an exchange of one defective helix for another. We thus propose that DNA glycosylases use a processive extrahelical scanning mechanism to locate nondestabilizing adducts (Figure 3b). We envisage that DNA, which initially associates with the glycosylase through electrostatic interactions, eventually acquires sufficient thermal energy to isomerize to an extrahelical structure in which a base is inserted into the enzyme active site. This structure is initially generated at a distance from the lesion. The enzyme then migrates along the DNA, placing each base along the strand into the active site in turn and sampling for complementarity. Eventually, the enzyme migrates to the site of the lesion, whereupon a Michaelis complex is assembled and catalysis ensues.

In summary, we propose that base-excision DNA repair proteins locate their substrates through either one of two distinct mechanisms, depending upon the propensity of the lesion to become spontaneously extrahelical. DNA-destabilizing lesions have an increased tendency to adopt an extrahelical conformation, and we propose that DNA glycosylases simply intercept the extrahelical form of these lesions. Lesions that cause little or no DNA destabilization require a more active mechanism of location by DNA glycosylases than do destabilizing lesions. In the latter case, we propose that the enzyme initially extrudes a base at a distance from the lesion; the enzyme and extrahelical base then migrate processively along the DNA, sampling each base in the enzyme active site until finally the lesion is reached, at which point catalysis ensues. These models should be readily testable through the introduction of synthetic probes into DNA. For example, placement of helical defects such as nicks, bulges and abasic site analogs [25,26,34] at a distance from a nondestabilizing lesion should accelerate the rate of base excision, whereas DNA-stiffening devices such as disulfide cross-links [35,36] placed at a distance should retard base excision.

Acknowledgements

We thank Huw Nash for insightful discussions and a critical reading of the manuscript. Published and ongoing X-ray crystallographic studies on AlkA are a collaborative effort with the laboratory of Tom Ellenberger (Harvard Medical School), to whom we are indebted for consultation. We gratefully acknowledge the National Institute of General Medical Sciences for supporting research on DNA glycosylases in the authors' laboratory (GM 51330).

References

1. Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715.
2. Friedberg, E.C., Walker, G.C. & Siede, W. (1995). *DNA Repair and Mutagenesis*, ASM Press Washington D.C., USA.
3. Wood, R.D. (1996). DNA repair in eukaryotes. *Annu. Rev. Biochem.* **65**, 135–167.
4. Nash, H.M., *et al.*, & Verdine, G.L. (1996). Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily. *Curr. Biol.* **6**, 968–980.
5. Van der Kemp, P.A., Thomas, D., Barbey, R., de Oliveira, R. & Boiteux, S. (1996). Cloning and expression in *Escherichia coli* of the OGG1 gene of *Saccharomyces cerevisiae*, which codes for a DNA glycosylase that excises 7,8-dihydro,8-oxoguanine and 2,6-diamino-4-hydroxy-5-N-methylformamidopyramidine. *Proc. Natl Acad. Sci. USA* **93**, 5197–5202.
6. Aspinwall, R., *et al.*, & Hickson, I.D. (1997). Cloning and characterization of a functional human homolog of *Escherichia coli* endonuclease III. *Proc. Natl Acad. Sci. USA* **94**, 109–114.
7. Hilbert, T.P., *et al.*, & Teebor, G.W. (1996). Purification of a mammalian homolog of *Escherichia coli* endonuclease III: Identification of a bovine pyrimidine hydrate-thymine glycol DNA-glycosylase/AP Lyase by irreversible cross linking to a thymine glycol-containing oligodeoxynucleotide. *Biochemistry*, **35**, 2505–2511.
8. Slupska, M.M., Baikalov, C., Luther, W.M., Chiang, J.H., Wei, Y.F. & Miller, J.H. (1996). Cloning and sequencing a human homolog (hMYH) of the *Escherichia coli* mutY gene whose function is required for the repair of oxidative DNA damage. *J. Bacteriol.* **178**, 3885–3892.
9. Lu, R., Nash, H.M. & Verdine, G.L. (1997). A mammalian DNA repair enzyme that excises oxidatively damaged guanines maps to a locus frequently lost in lung cancer *Curr. Biol.* **7**, 397–407.
10. Labahn, J., Scharer, O.D., Long, A., Ezaz-Nikpay, K., Verdine, G.L. & Ellenberger, T.E. (1996). Structural basis for the excision repair of alkylation-damaged DNA. *Cell* **86**, 321–329.
11. Yamagata, Y., *et al.*, & Fujii, S. (1996). Three-dimensional structure of a DNA repair enzyme, 3-methyladenine DNA glycosylase II, from *Escherichia coli*. Cell **86**, 311–319.
12. Dodson, M.L., Schrock, R.D. III. & Lloyd, R.S.(1993). Evidence for an imino intermediate in the T4 endonuclease V reaction. *Biochemistry* **32**, 8284–8290.
13. Sun, B., Latham, K.A., Dodson, M.L. & Lloyd, R.S. (1995). Studies on the catalytic mechanism of five DNA glycosylases. Probing for enzyme-DNA imino intermediates. *J. Biol. Chem.* **270**, 19501–19508.
14. Zharkov, D.O., Rieger, R.A., Iden, C.R. & Grollman, A.P. (1997). NH_2-terminal proline acts as a nucleophile in the glycosylase/AP-lyase reaction catalyzed by *Escherichia coli* formamidopyrimidine-DNA glycosylase (Fpg) protein. *J. Biol. Chem.* **272**, 5335–5341.
15. Thayer, M.M., Ahern, H., Xing, D.X., Cunningham, R.P. & Tainer, J.A. (1995). Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. *EMBO J.* **14**, 4108–4120.
16. Kuo, C-F., McRee, D.E., Fisher, C.L., O'Handley, S.F., Cunningham, R.P. & Tainer, J.A. (1992). Atomic structure of the DNA repair [4Fe-4S] enzyme endonuclease III. *Science* **258**, 434–440.
17. Doherty, A.J., Serpell, L.C. & 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.

18. Verdine, G.L. (1994). The flip side of DNA methylation. *Cell* **76**, 197–200.
19. Klimasauskas, S., Kumar, S., Roberts, R.J. & Cheng, X. (1994). Hha I methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357–369.
20. 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.
21. Slupphaug, G., Mol, C.D., Kavli, 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.
22. Vassylyev, D.G., *et al.*, & 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.
23. Kao, J.Y., Goljer, I., Phan, T.A. & Bolton, P.H. (1993). Characterization of the effects of a thymine glycol residue on the structure, dynamics, and stability of duplex DNA by NMR. *J. Biol. Chem.* **268**, 17787–17793.
24. Morden, K.M., Gunn, B.M.,& Maskos, K. (1990). NMR Studies of a deoxyribodecanucleotide containing an extrahelical thymidine surrounded by an oligo(dA)•oligo(dT) tract. *Biochemistry* **29**, 8835–8845.
25. Castaing, B., Boiteux, S. & Zelwer, C. (1992). DNA containing a chemically reduced apurinic site is a high affinity ligand for the *E. coli* formamidopyrimidine-DNA glycosylase. *Nucleic Acids Res.* **20**, 389–394.
26. Schärer, O.D. & Verdine, G.L. (1995). A designed inhibitor of base-excision DNA repair. *J. Am. Chem. Soc.* **117**, 10781–10782.
27. McCullough, A.K., Schärer, O.D., Verdine, G.L. & Lloyd, R.S. (1996). Structural determinants for specific recognition by T4 endonuclease V. *J. Biol. Chem.* **271**, 32147–32152.
28. von Hippel, P.H. & Berg, O.G. (1989). Facilitated target location in biological systems. *J. Biol. Chem.* **264**, 675–678.
29. Ezaz-Nikpay, K. & Verdine, G.L. (1992). Aberrantly methylated DNA: site-specific insertion of N7-methyl-2′-deoxyguanosine into the Dickerson/Drew dodecamer. *J. Am. Chem. Soc.* **114**, 6562–6563.
30. Ezaz-Nikpay, K. & Verdine, G.L. (1994). The effect of N7-methylguanine on duplex DNA structure. *Chem. Biol.* **1**, 235–240.
31. Plum, E.G., Grollman, A.P., Johnson, F. & Breslauer, K.J. (1995). Influence of the oxidatively damaged adduct 8-oxodeoxyguanosine an the conformation, energetics, and thermodynamic stability of a DNA duplex. *Biochemistry* **34**, 16148–16160.
32. Moe, J. & Russu, I. (1992). Kinetics and energetics of base-pair opening in 5′-d(CGCGAATTCGCG)-3′ and a substituted dodescamer containing G•T mismatches. *Biochemistry* **31**, 8421–8428.
33. Wu, J.C. & Santi, D.V. (1987). Kinetic and catalytic mechanism of HhaI methyltransferase. *J. Biol. Chem.* **262**, 4778–4786.
34. Takeshita, M., Chang, C-N., Johnson, F., Will, S. & Grollman, A.P. (1987). Oligodeoxynucleotides containing synthetic abasic sites. *J. Biol. Chem.* **262**, 10171–10179.
35. Ferentz, A.E. & Verdine, G.L. (1991). Disulfide cross-linked oligonucleotides. *J. Am. Chem. Soc.* **113**, 4000–4002.
36. Ferentz, A.E. & Verdine, G.L. (1994). The convertible nucleoside approach: structural engineering of nucleic acids by disulfide cross-linking. In *Nucleic Acids and Molecular Biology* (Eckstein, F. & Lilley, D.M.J., eds) pp.14–40, Springer-Verlag New York, USA.