

Binding of Specific DNA Base-pair Mismatches by *N*-Methylpurine-DNA Glycosylase and Its Implication in Initial Damage Recognition

Tapan Biswas¹, Lawrence J. Clos II², John SantaLucia Jr²
Sankar Mitra¹ and Rabindra Roy^{1*}

¹Sealy Center for Molecular Science and Department of Human Biological Chemistry and Genetics, University of Texas Medical Branch Galveston TX 77555-1079, USA

²Department of Chemistry Wayne State University Detroit, MI 48202, USA

Most DNA glycosylases including *N*-methylpurine-DNA glycosylase (MPG), which initiate DNA base excision repair, have a wide substrate range of damaged or altered bases in duplex DNA. In contrast, uracil-DNA glycosylase (UDG) is specific for uracil and excises it from both single-stranded and duplex DNAs. Here we show by DNA footprinting analysis that MPG, but not UDG, bound to base-pair mismatches especially to less stable pyrimidine–pyrimidine pairs, without catalyzing detectable base cleavage. Thermal denaturation studies of these normal and damaged (e.g. 1,*N*⁶-ethenoadenine, ϵ A) base mispairs indicate that duplex instability rather than exact fit of the flipped out damaged base in the catalytic pocket is a major determinant in the initial recognition of damage by MPG. Finally, based on our determination of binding affinity and catalytic efficiency we conclude that the initial recognition of substrate base lesions by MPG is dependent on the ease of flipping of the base from unstable pairs to a flexible catalytic pocket.

© 2002 Elsevier Science Ltd. All rights reserved

Keywords: *N*-methylpurine-DNA glycosylase; mismatch; DNA instability; substrate recognition; catalytic efficiency

*Corresponding author

Introduction

Repair of DNA containing small DNA adducts, as well as altered and abnormal bases, occurs primarily *via* the base excision repair (BER) pathway, whose first step is cleavage of the base by a DNA glycosylase in all organisms.^{1–2} Accumulated evidence on the properties of many DNA glycosylases indicates the presence of two classes of DNA glycosylases on the basis of their substrate specificity.^{2–3} UDG from both *Escherichia coli* and mammals, which is the most extensively studied glycosylase and is often used as the paradigm for the DNA glycosylase reaction, constitutes a distinct class that acts nearly exclusively on uracil in

DNA.^{4–5} It excises U from G·U mismatch or A·U equally well, is active on both single and double-stranded DNA substrates, and in fact prefers single-stranded DNA.^{6–7} In contrast, other DNA glycosylases such as mammalian MPG, *E. coli* endonuclease III (Nth), and formamido pyrimidine-DNA glycosylase³ (Fpg or MutM), which have a broad specificity for substrates with widely different structures, constitute the second class. The hallmark of this second group is low turnover, possibly several orders of magnitude lower than that of UDG,^{7–10} and extremely low activity if any with single-stranded DNA substrates. How this second group of enzymes recognizes their cognate substrate lesions in DNA is an important question because in spite of their broad substrate range, the enzymes do not act indiscriminately on any DNA lesion.

The mammalian MPG is known to excise at least 17 structurally diverse modified bases from DNA.¹¹ These lesions include 3-alkylpurines, 7-alkyl-guanine, 1,*N*⁶-ethenoadenine, *N*²,3-ethenoguanine, and hypoxanthine (Hx), all of which are purine derivatives.^{8,12–16} Moreover, the base alterations are located in both the major and minor grooves

Present addresses: T. Biswas, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA; R. Roy, Division of Carcinogenesis and Molecular Epidemiology, American Health Foundation, Valhalla, NY 10595, USA.

Abbreviations used: MPG, *N*-methylpurine-DNA glycosylase; UDG, uracil-DNA glycosylase; ϵ A, 1,*N*⁶-ethenoadenine; Hx, hypoxanthine.

E-mail address of the corresponding author: roy@ahf.org

```

FA 5' TCGAGGATCCTGAGCTCGAGTCGACGA TCGCGAATTCTGCGGATCCAAGC3'
FT 5' TCGAGGATCCTGAGCTCGAGTCGACGT TCGCGAATTCTGCGGATCCAAGC3'
FG 5' TCGAGGATCCTGAGCTCGAGTCGACGG TCGCGAATTCTGCGGATCCAAGC3'
FC 5' TCGAGGATCCTGAGCTCGAGTCGACGC TCGCGAATTCTGCGGATCCAAGC3'
FU 5' TCGAGGATCCTGAGCTCGAGTCGACGU TCGCGAATTCTGCGGATCCAAGC3'
FEA 5' TCGAGGATCCTGAGCTCGAGTCGACGA TCGCGAATTCTGCGGATCCAAGC3'

RA 3' AGCTCCTAGGACTCGAGCTCAGCTGCAAGCGCTTAAGACGCCTAGGTTCCG5'
RT 3' AGCTCCTAGGACTCGAGCTCAGCTGCTAGCGCTTAAGACGCCTAGGTTCCG5'
RG 3' AGCTCCTAGGACTCGAGCTCAGCTGCGAGCGCTTAAGACGCCTAGGTTCCG5'
RC 3' AGCTCCTAGGACTCGAGCTCAGCTGCCAGCGCTTAAGACGCCTAGGTTCCG5'

```

Figure 1. Sequence of oligonucleotide for normal and mismatched duplex DNAs. F, forward strand and R, reverse strand.

of duplex DNA. Its orthologs in *E. coli* (AlkA) and in yeast (MAG) have an overlapping although not identical substrate range. But in spite of this functional similarity, mammalian MPG and *E. coli* AlkA do not share significant sequence similarity or structural homology^{17–18} even when 3-methyladenine is a preferred substrate for all of them. MPG excises ϵ A and Hx more efficiently than AlkA and MAG,¹⁵ but unlike AlkA, it cannot excise O^2 -alkylpyrimidines^{19–20} and oxidized bases, such as 5-formyluracil and 5-hydroxymethyluracil²¹ from DNA. MAG also does not excise O^2 -methylthymine.^{11,22}

While the tertiary structures of many DNA glycosylases, e.g. UDG, AlkA, MPG, Nth, OGG1, and mismatch-specific uracil DNA-glycosylase (MUG) have recently been elucidated by X-ray crystallography, the initial steps in their substrate recognition are still not clear. In the case of U excision by UDG, a general mechanism has been postulated which involves a series of steps initiated by active “flipping” of the nucleotide containing the substrate base from the helix followed by tight fitting of the base in the catalytic pocket. Protonation of the base followed by cleavage of the glycosylic bond, presumably with activated water, completes the reaction.^{5,23–26} Obviously, this nearly exact fit of the flipped-out base in the catalytic pocket cannot be extended to AlkA and other DNA glycosylases, which recognize substrates of diverse structures. Based on the structure, it has been proposed that AlkA recognizes electron-deficient methylated bases through pi-electron donor/acceptor interactions involving the electron-rich aromatic cleft.¹⁷ However, this hypothesis did not include other considerably larger and/or non-electron-deficient substrates, e.g. 1,*N*⁶-etheno-adenine (ϵ A) and hypoxanthine (Hx). Recent elucidation of the crystal structure of a complex of MPG with DNA containing ϵ A, suggested that after the base is flipped-out, accommodation of the damaged base, as opposed to the normal base, in the substrate-binding pocket depends on its shape, hydrogen-bonding characteristics, and aromaticity.²⁷ In any case, the prerequisite to flipping, active site accommodation and catalysis is expected to be the initial recognition of the unpaired or abnormally paired bases by the enzyme at the damaged site. Assuming that DNA

glycosylases recognize various substrates by scanning and pinching the DNA backbone,^{28–30} it is important to identify the common determinant(s) in the substrate DNA that initially attracts the enzyme to the damage site for primary binding, and the features that finally determine the specificity of cleavable bases in the catalytic pocket. In fact, it is postulated in a recent report that MPG may initially recognize the structural distortions in DNA induced by the damage as a basis for mechanism of recognition of substrates of widely different structures.³¹ Our data indicate that MPG, unlike UDG, recognizes DNA distortions at specific normal base-pair mismatches without catalyzing the subsequent cleavage reaction. These results, correlated with thermodynamic stability of base-pairing and catalytic efficiency of excision of ϵ A, have led to a model for initial substrate recognition by MPG.

Results

MPG binds mismatched base-pairs in DNA

While investigating the interaction of MPG with substrate base lesions in DNA, we serendipitously discovered that the enzyme binds to mismatched pairs of normal bases. DNase I footprinting studies for evaluating binding of MPG to eight such possible mismatched base-pairs in oligonucleotides (Figure 1) show that mouse MPG stably binds selectively to the pyrimidine-containing single base-pair mismatches T·T, C·T and C·C (Figure 2(a) and (b)). The binding to the purine–pyrimidine mismatches, i.e. A·C and T·G pairs, was barely detectable under identical experimental conditions (Figure 2(a), lanes 7–9 and 13–15) and purine–purine mismatches did not show any detectable binding. As expected, oligonucleotides with the normal Watson–Crick base-pairs gave no footprint at all (up to 0.8 μ M MPG; Figure 2(a), lanes 22–24 and Figure 3(a), lanes 1–3). Surprisingly, MPG appears to have a much higher affinity for the U·T (Figure 3(a)) or U·C mismatch (data not shown) than for other base-pair mismatches. The apparent dissociation constant (K_d^{app}) as determined from quantitative analysis of the extent of protection from DNase I was calculated to be about 21 nM

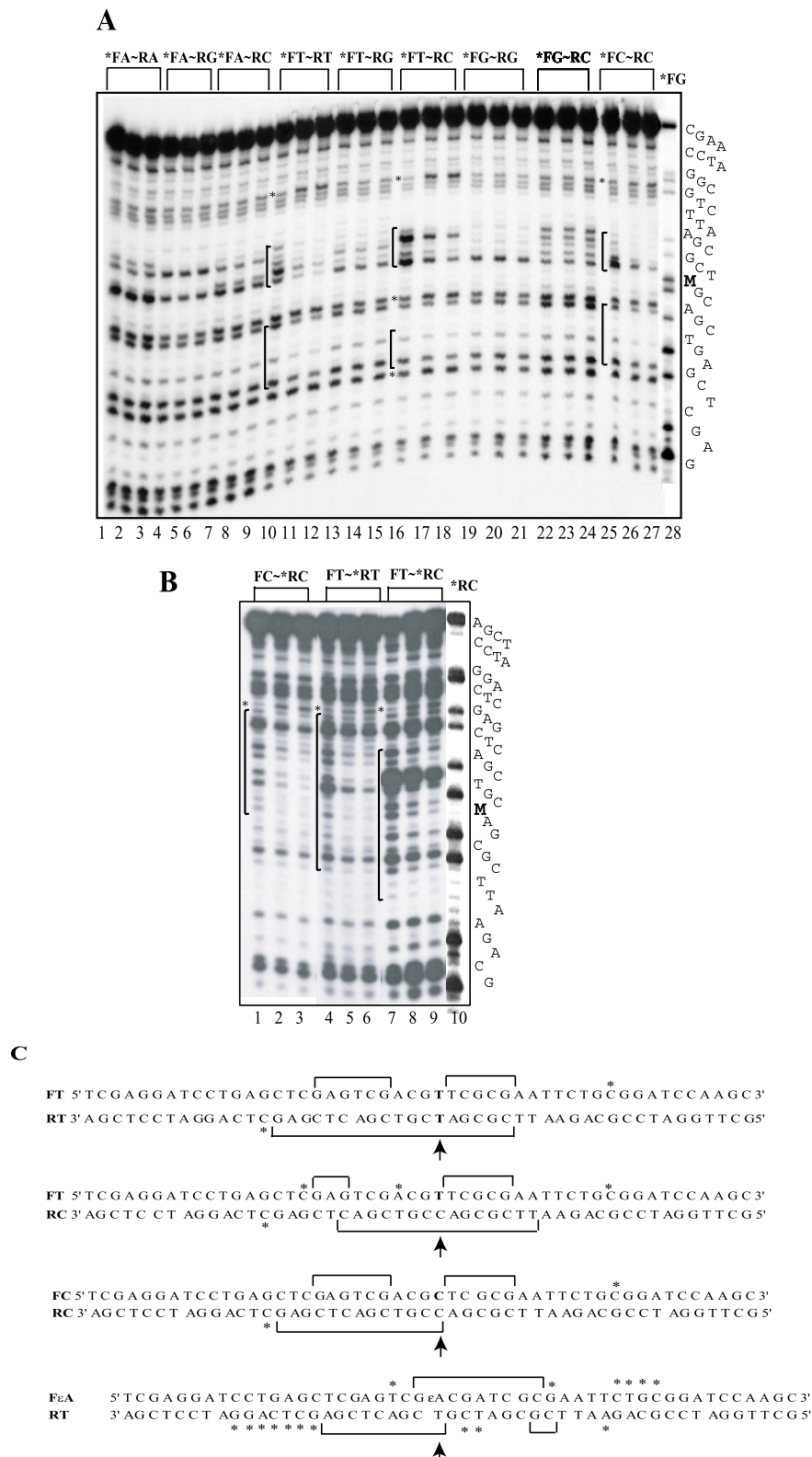


Figure 2. DNase I footprinting analysis of mismatched base-pair binding by MPG. (a) The forward strands of base-pair mismatch (lanes 1–21, 25–27) and control (lanes 22–24) containing oligonucleotide duplexes (as shown above each triplet of lanes) were incubated in the presence of 0 (lanes 1, 4, 7, 10, 13, 16, 19, 22, 25), 20 (lanes 2, 5, 8, 11, 14, 17, 20, 23, 26), or 40 (lanes 3, 6, 9, 12, 15, 18, 21, 24, 27) pmol of MPG and subjected to DNase I footprinting as described in Materials and Methods. Lane 28, Maxam–Gilbert G sequencing reaction on the forward strand of FG ~ RC. M indicates the site of mismatch in the duplex. (b) The reverse strands of mismatches (lanes 1 to 9) which bound to MPG in (a) were incubated with 0, 20, and 40 pmol of MPG and subsequently analyzed as described in (a). Lane 10, Maxam–Gilbert G sequencing reaction on the reverse strand of FT ~ RC. (c) Protected regions are marked by brackets and hypersensitive sites by asterisks. The DNase I footprinting profile of FeA ~ RT is included from our previous study.⁸

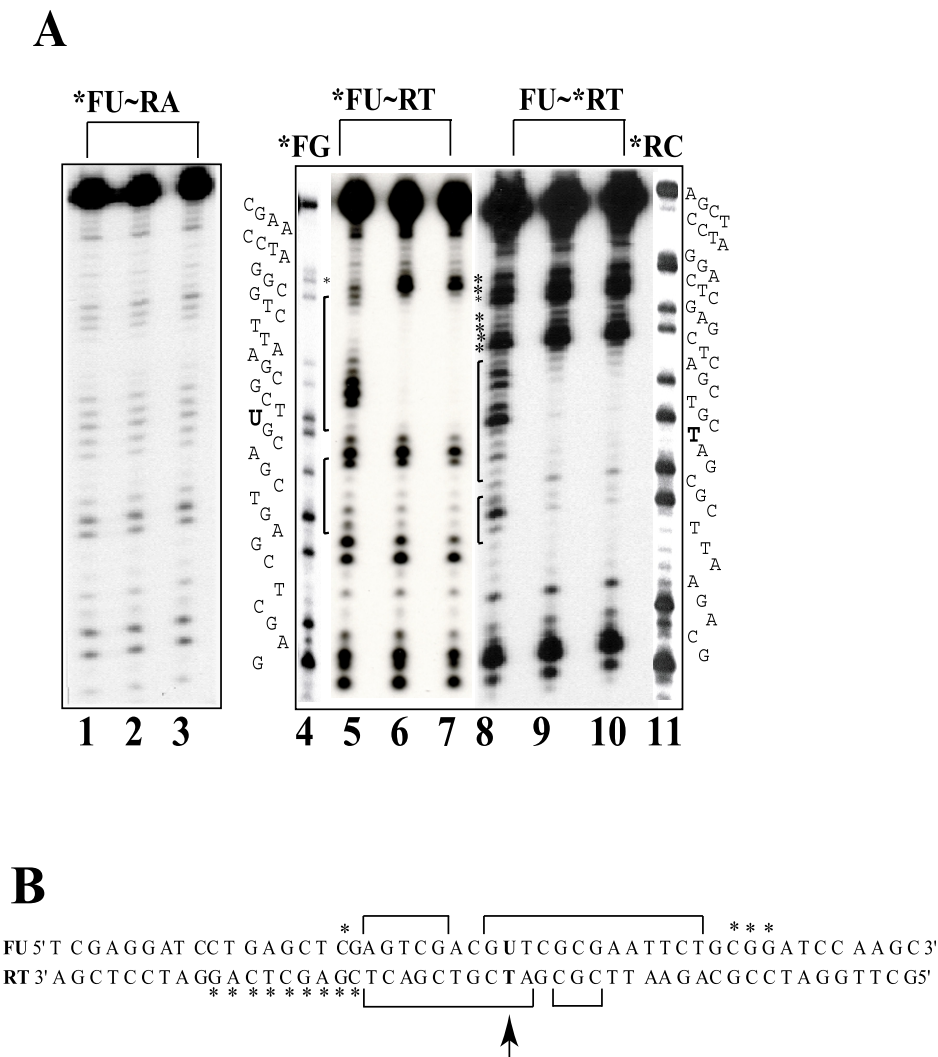


Figure 3. DNase I footprinting of U·T mismatch-containing duplex oligonucleotide complexed with MPG. A 50-mer oligonucleotide containing a U·T mismatch was incubated \pm MPG and subjected to DNase I footprinting. (a) Lanes 1–3, DNase I digestion pattern of the forward strand in the DNA duplex containing U·A pair in the presence of 0, 10 and 20 pmol of MPG, respectively. Lanes 5–7 and 8–10 show the protection patterns in the forward and reverse strands containing U opposite T, respectively, with 0, 10 and 20 pmol of MPG. Lanes 4 and 11 represent the Maxam–Gilbert G sequencing reaction on the forward and reverse strands of FG ~ RC, respectively. (b) Protected regions are marked by brackets and hypersensitive sites by asterisks.

for the T·T and 1.5 nM for the U·T pair, compared to the K_d^{APP} value of 0.2 nM for the substrate ϵ A·T pair (Table 1). In all cases studied the mismatch binding by MPG was catalytically non-productive.

DNase I footprinting analyses indicate that MPG forms a specific binding complex that generates a distinct pattern of protection spanning a region of 15 nucleotides on the forward strand and 13–18 nucleotides on the reverse strand containing T·T, T·C and C·C mismatches in a 50 bp oligonucleotide duplex (Figure 2). In contrast, in U·T mismatch containing a duplex of identical sequence, strong protection of a region spanning 20 nucleotides on the forward strand and 14 nucleotides on the reverse strand was observed (Figure 3(b)). Increased DNase I sensitivity of certain nucleotides located outside the protected region due to protein

binding is a common feature of many DNase I footprinting profiles which is believed to be a signature of DNA distortions. We have included the footprinting profile of the ϵ A·T duplex shown in Figure 2(c) from an our earlier study for comparison.⁸ Even though DNase I footprinting is a good indicator of differences in binding of MPG to damaged or mismatched base-pairs, meaningful comparison of footprinting profiles is not possible because of low resolution and sequence-specific cleavage.

Lack of mismatch binding by UDG

We carried out similar DNase I footprinting studies of the same mismatched DNA duplexes as used earlier, with human UDG. UDG provided no

Table 1. Binding affinities and initial velocities for cleavage of MPG for 50-mer duplex oligonucleotides containing ϵ A opposite A, T, G, C and AP site, U opposite T and T opposite T

Base-pair	K_d^{app} (nM)	ϵ A excised (nM/min)
ϵ A-A	0.76	1.5
ϵ A-T	0.2	2.0
ϵ A-G	2.0	0.16
ϵ A-C	0.2	2.5
ϵ A-AP site	2.6	0.4
T-T	21.0	ND
U-T	1.5	ND

Binding affinity was measured from quantitative DNase I footprinting analysis. See Materials and Methods for details. Initial velocity was calculated from the linear part of the reaction shown in Figure 5, using 10 nM MPG and 100 nM substrate oligonucleotides. See Materials and Methods for details. ND, not detectable.

detectable protection to the oligonucleotide duplex containing any of the eight normal base mismatches (data not shown). We could not determine the substrate binding of UDG using the U-containing oligonucleotide by a similar footprinting assay. Because of high turnover rate of this enzyme, compared to other DNA glycosylases, substrates were consumed almost instantaneously. Since these monofunctional glycosylases do not require any cofactor for their activity, it is difficult to prevent or slow down their reaction.

Recognition and catalytic activity of MPG on ϵ A opposite A, T, G, C and abasic (AP)-site

In an effort to evaluate the contribution of the opposite base-to-substrate recognition and catalytic activity, we determined the binding affinity of MPG for ϵ A incorporated in a 50-mer oligonucleotide duplex (of an identical sequence used earlier) in which the base opposite ϵ A was either A, T, G,

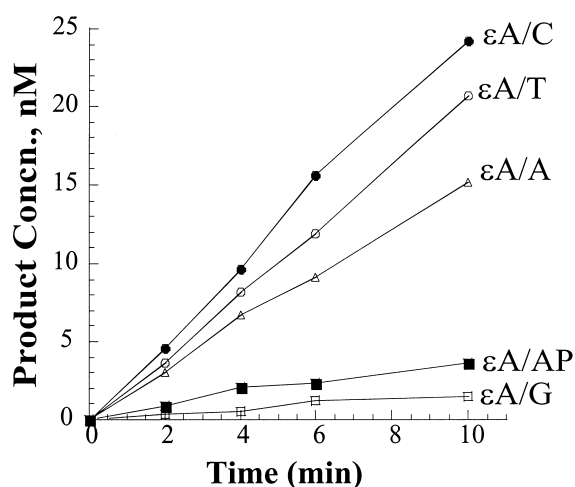


Figure 4. Kinetics of release of ϵ A when paired with A, T, G, C or AP-site in a 50-mer duplex oligonucleotide. The details of the assay are described in Materials and Methods.

C or an AP-site (Table 1). At the same time, we measured the initial velocity of MPG for excision of ϵ A using the same oligonucleotide duplexes containing various bases opposite ϵ A (Figure 4 and Table 1). The strongest binding and highest catalytic activity was observed when the lesion is present opposite T or C. The presence of A opposite ϵ A reduced both binding and catalytic activity. Little activity was detected when ϵ A was positioned opposite G or an AP-site and interestingly, these also showed very weak binding affinity similar to that observed with the U-T mismatch pair.

Determination of stability of DNA duplexes

In order to test the notion that MPG recognizes initially the structural deformation in DNA induced by both mispaired normal and damaged bases, we compared their destabilizing effect on the duplex DNA. Eighteen nucleotide long duplex DNAs, containing the mismatches in the same sequence context as used in DNA footprinting and excision activity studies, were used. Analysis using the MFold program predicted no significant secondary structure formation for any of the strands being melted except the A-complement. An experimental melting study of the A-complement strand, however, showed no indication of secondary structure formation (data not shown). All further calculations were made assuming a two-state model.

For each duplex, eight to ten melting curves were generated at concentrations ranging from 50 to 100-fold and the standard state enthalpy and entropy changes. ΔH° and ΔS° values were calculated using the program MELTWIN.³² Using these values, the free-energy change, ΔG_{37° , and melting temperature, T_m , were calculated.³³ The curve fitted data calculated in this way were averaged for each sequence. No systematic temperature dependence of the fitted ΔH° value was observed for any of the sequences, indicating that ΔC_p° is close to zero. A plot of reciprocal T_m versus $\ln(C_t/4)$ was used to independently derive thermodynamic parameters.³⁴

Comparison of the ΔH° data, derived from the $1/T_m$ versus $\ln(C_t/4)$ plot, with the averaged curve fitted data showed that in all cases two calculated ΔH° values agreed within 15%. This suggests that the two-state approximation is valid for all duplexes.³⁵ The error-weighted averages of the data from the two different methods were calculated,³⁶ and are presented in Table 2. The stability of ϵ A pairing shows the trend $G > A > T \approx C \approx U > AP$.

The melting temperatures of DNA duplexes containing normal base mismatches were computed by the HYTHER™ server. Duplexes containing pyrimidine-pyrimidine pairs (T-T, T-C, C-C) have considerably lower melting temperatures compared to the normal duplexes with A-T

Table 2. Thermodynamic parameters of ϵ A pairing

Base-pair	ΔH° (kcal/mol)	ΔS° (cal/K mol)	ΔG_{37}° (kcal/mol)	T_m ($^\circ$ C) ($C_T = 10^{-4}$ M)
ϵ A-A	-117.2 ± 2.6	-325.2 ± 7.6	-16.3 ± 0.2	65.2
ϵ A-T	-122.7 ± 4.1	-343.2 ± 12.4	-16.3 ± 0.3	63.9
ϵ A-G	-120.0 ± 4.9	-330.1 ± 14.4	-17.5 ± 0.5	67.9
ϵ A-C	-122.5 ± 4.2	-342.9 ± 12.5	-16.2 ± 0.3	63.5
ϵ A-U	-117.6 ± 2.0	-327.9 ± 6.0	-15.9 ± 0.2	63.5
ϵ A-AP	-113.8 ± 2.0	-316.9 ± 5.6	-15.6 ± 0.2	63.4

See Materials and Methods for sequence context and solution conditions.

or G-C or purine-purine and purine-pyrimidine pairs (Table 3).

Discussion

Our DNA footprinting analysis shows that MPG binds to pyrimidine-pyrimidine base mismatches quite efficiently. A common structural feature of these pyrimidine-pyrimidine mismatches in duplex DNA is their significant flexibility and thermodynamic instability, unlike that of the purine-pyrimidine or purine-purine base-pairs.³⁷ Our results indicate that this instability of base-pairs is the common denominator, not only for the normal base-pair mismatches that are recognized by MPG but for the damage-containing mismatches as well.

The instability of specific base-pairs is a consequence of cumulative effects of pairing with opposite and stacking with neighboring bases. We determined experimentally the T_m values of various oligonucleotide duplexes having identical sequence except for a single base-pair mismatch of normal bases or of an ϵ A lesion opposite any normal base or an abasic site. The difference in T_m value is thus a quantitative measure of the difference in stability or the structural integrity of the duplex substrate arising out of the mismatched pair alone, and eliminates the potential role of the DNA sequence context.³⁸ We also measured the K_d^{app} value of MPG for various duplex DNA substrates by using quantitative DNA footprinting and re-examined the relation of base-pair instability to substrate binding and activity of the enzyme. The ϵ A-G pair, which forms Hoogsteen H-bonding,³⁹⁻⁴⁰ was found to have the highest T_m

value among the pairs. It also has the highest relative K_d^{app} value and the lowest catalytic efficiency of ϵ A cleavage (determined in this study) among ϵ A pairs with normal bases. T and C, and to a lesser extent A, when present opposite ϵ A, showed lower stability (i.e. lower T_m value) and show both better binding and excision activity of MPG. In all cases, the K_d^{app} values correlated well with the relative base excision activity in duplex oligonucleotides containing ϵ A. A simple explanation of poor substrate activity thus seems to be related to the difficulty of flipping ϵ A to an extra-helical state when base-pairing is stronger. This inverse correlation between T_m value and activity supports the idea of early recognition of an unstable duplex site, which could determine rates for multi-substrate enzymes like MPG with low turnover. However, this cannot explain the anomaly of extremely weak activity and inefficient binding of MPG to the least stable ϵ A-AP pair (Table 1). This paradox could be resolved if MPG requires a complementary base in duplex DNA opposite the lesion. While instability of the lesion base-pair enhances enzyme activity by promoting substrate flipping, the opposite base which is missing in the ϵ A-AP site pair could play an important role in stabilizing the enzyme substrate complex at an initial step of binding prior to catalysis. It is also surprising that even though MPG binds to an ϵ A-A pair and excises the ϵ A residue with relatively high efficiency, it binds quite poorly to an A-A mismatch. This could be due to inability of MPG to accommodate A in the catalytic pocket or/and A as an unsuitable opposite base. Another interesting observation was about stronger binding of MPG to a U-T than T-T mismatch in spite of their

Table 3. Predicted thermodynamic parameters of normal and mismatch pairing

Base-pair	ΔH° (kcal/mol)	ΔS° (cal/K mol)	ΔG_{37}° (kcal/mol)	T_m ($^\circ$ C) ($C_T = 10^{-4}$ M)
A-T	-148.2	-416.8	-18.9	67.3
G-C	-149.2	-416.5	-20.0	67.8
A-A	-134.5	-382.2	-16.0	60.4
A-G	-134.1	-377.6	-17.0	63.2
A-C	-122.3	-345.4	-15.2	60.6
T-T	-137.6	-393.4	-15.6	58.8
T-C	-126.9	-360.3	-15.1	59.6
T-G	-136.2	-385.5	-16.6	61.9
G-G	-141.9	-399.5	-18.0	64.3
C-C	-129.3	-369.6	-14.6	57.8
U-T	-137.6	-393.4	-15.6	58.8

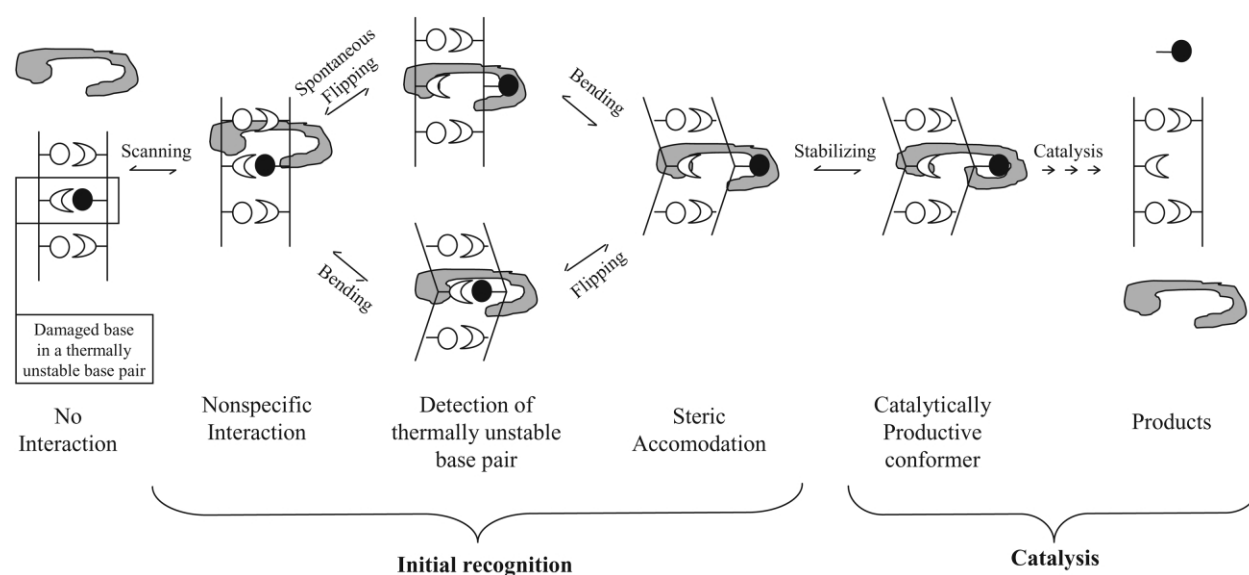


Figure 5. Proposed model of initial substrate recognition and catalysis by MPG. The details are provided in the text.

similar effect on duplex destabilization (reflected in similar T_m value). It is possible that U fits better than T in the active site or that it is a better opposite base. The absence of a methyl group in U is a possible reason for its easier accommodation in the active site. In fact a co-crystal structure of MPG indicates that in flipped-out T, the methyl group extends towards the surface of the active site pocket. On the other hand, because we did not observe any noticeable difference in binding or activity of MPG towards an ϵ A·T *versus* ϵ A·U base-pairs (data not shown), it is unlikely that U acts as a better opposite base. However, a recent report indicates that even though MPG has similar substrate specificity for ϵ A·T and ϵ A·U, the k_D value for ϵ A·U appeared to be threefold higher.³¹ These results therefore warrant a re-examination of the role of the opposite base which may determine not only the instability of the base-pair, facilitating nucleotide-flipping, but which also may interact specifically with the enzyme. Although an earlier study did not observe a similarly large effect of the opposite base on the reaction rate of MPG,⁴¹ it is possible some unusual neighboring sequence or size of the DNA substrate used in that study might have modulated this effect. Other previous studies have shown that the reaction rates of some DNA glycosylases in cleavage of a damaged or abnormal substrate base from duplex DNA vary significantly depending upon the nature of the opposite base. The mismatch-specific uracil-DNA glycosylase (MUG), which unlike UDG (*E. coli* UNG and its human homologue) is active only on duplex DNA, provides an extreme example. Here the opposite base G of the G·U mismatch pair substrate interacts stably with the enzyme, and plays a critical role in substrate recognition.⁴² The situation of UDG displays a sharp contrast. It is unique among DNA glycosylases in having a single substrate. It is also the only

DNA glycosylase to be highly active on single-stranded DNA, and is barely affected by either base-pair stability or the nature of the opposite base. It excises U from G·U and A·U equally well⁷ as described in one study although a difference was reported in another.⁴³ Our inability to observe binding of UDG to any base-pair mismatch suggests that unusual base-pairing is not required for UDG's substrate recognition and the observation conforms well with the lack of requirement of a duplex DNA substrate.

It is somewhat unexpected that mismatched pyrimidine bases are recognized selectively by MPG. MutS, with no DNA base excision activity, binds to mismatched DNA base-pairs as a prerequisite to their repair. However, it binds to different mismatches with variable efficiency, and in fact binds poorly to the pyrimidine–pyrimidine mismatches.^{44–45} These observations suggest that the initial steps in the recognition of substrates in base excision repair and mismatch repair processes are distinct. It is likely that the primary determinant for the initial recognition of MPG for its substrate is a flexible (thermodynamically unstable) site in the DNA duplex whereas such flexibility could be responsible for low affinity for MutS.⁴⁶ Increased flexibility in the DNA duplex due to the presence of damage sites, e.g. urea, thymine glycol, and abasic site, has been proposed to be the basis for substrate recognition by other multi-substrate enzymes including endonuclease III and AP-endonuclease.⁴⁷ Flexibility due to disruption of base-pairs and changes in sugar–phosphate backbone structure of DNA has also been implicated in multi-substrate recognition of the nucleotide excision repair process,⁴⁸ whose first step could be similar to that of the base excision repair process.

Based on the observations reported here and of other studies, we have outlined a model for substrate recognition by MPG (Figure 5). In this

model, MPG scans or remains associated with the DNA (non-specific binding) predominantly *via* electrostatic interactions. A flexible (thermodynamically unstable) site of mismatched or damaged base-pairs has significantly higher affinity for MPG probably because of pronounced bending and/or nucleotide flipping. Accommodation of the base-pair (depending on the steric constraints of the flipped-out and/or the opposite base) in the catalytic pocket determines initial recognition and binding. Subtle conformational changes at DNA damage sites combined with the inherent plasticity of the catalytic pocket of MPG may explain its multiple substrate specificity. However, following initial binding, catalysis will occur only if the nucleotide is stabilized in the catalytic pocket for direct, in-line displacement. The lack of glycosylase activity on the pyrimidine pairs in spite of their moderately strong binding to MPG could be explained by the absence of appropriate interactions to stabilize the flipped out base from a pyrimidine pair in the catalytic pocket, because it may not be positioned appropriately for its protonation and/or nucleophilic attack on the C1' residue of the deoxynucleoside. Thus, there are multiple stages of active interaction between the protein and the DNA where a damaged base in an unnatural duplex is differentiated from a normal base in a Watson-Crick base-pair, giving rise to the final specificity which could be investigated by X-ray crystallographic studies of DNA glycosylases bound to distinct substrates and mismatched duplexes.

Materials and Methods

Purification of recombinant mouse MPG

The mouse MPG deletion mutant (NΔ100CΔ18) was over expressed in *E. coli* and purified as described and used throughout this study.⁸

DNase I footprinting analysis

Oligonucleotides (F; 50-mer) containing a single variable residue (U, A, T, G and C) at position 27 and complementary sequence (R; 50-mer) also containing a single variable residue A, T, G and C at position 24 were synthesized using standard phosphoramidite chemistry. The sequences of the oligonucleotide primers are shown in Figure 1. The single-stranded oligonucleotides were purified on a denaturing 10% (w/v) polyacrylamide gel before 5' end-labeling with T4 polynucleotide kinase and [γ -³²P]ATP. Complementary strands were then annealed in TEN buffer (10 mM Tris-HCl (pH 7.8), 1 mM EDTA, 100 mM NaCl) and the duplex oligonucleotide was purified through a G-50 column to remove unincorporated nucleotides. DNase I footprinting was performed essentially as described by Leblanc & Moss.⁴⁹ The binding reaction (50 μ l) was performed with 4 fmol of ³²P end-labeled oligonucleotide (~15,000 cpm), 50 ng of plasmid DNA, and varying amounts of NΔ100CΔ18 MPG and UDG in a buffer containing 35 mM Hepes-KOH (pH 7.8), 0.5 mM

EDTA, 0.5 mM DTT, 60 mM KCl, and 10% (v/v) glycerol. After incubating for five minutes at room temperature, the DNA was digested with 5 μ l of 0.002 unit (Kunitz)/ μ l DNase I (Sigma) for two minutes. The reaction was stopped by addition of reaction stop buffer (1% (w/v) SDS, 200 mM NaCl, 20 mM EDTA (pH 8.0), and 40 μ g/ml tRNA), and after extraction with phenol/chloroform, the DNA was precipitated in ethanol, dried, dissolved in loading buffer containing 66% (v/v) formamide and 0.03 M NaOH, and then resolved by electrophoresis on a sequencing gel containing 8% polyacrylamide/7 M urea.

Determination of binding affinities

Two independent major DNase I protected bands in the ϵ A-containing strand protected from DNase I, in the presence of MPG, were selected for quantification of radioactivity in the PhosphorImager (Molecular Dynamics, Sunnyvale, CA). The MPG concentration that reduced the radioactive counts in the bands of interest to 50% was considered as the K_d^{APP} value.

Determination of catalytic activity of MPG

MPG (10 nM) was incubated with the ϵ A-containing duplex oligonucleotide substrate (50-mer, 100 nM) for zero to ten minutes at 37 °C in the MPG assay buffer (20 μ l) containing 25 mM Hepes-KOH (pH 7.8), 0.5 mM EDTA, 0.5 mM DTT, 5% glycerol and 150 mM NaCl. Duplex containing mismatches were generated by mixing equimolar amounts of complementary strands and then annealing by raising the temperature to 99 °C and cooling it to 4 °C at a rate of 0.01 deg. C/second in 10 mM Tris-HCl (pH 7), 1 mM EDTA, and 100 mM NaCl. The reaction was terminated at different times by directly adding 20 μ l of 2 M piperidine and then incubating at 95 °C for 25 minutes. DNA was precipitated with ethanol in the presence of 0.3 M sodium acetate and resolved on an 8% polyacrylamide sequencing gel. The radioactivity in the bands (the uncleaved DNA of 50-mer and the cleaved product of 26-mer) of the gel was quantified with a PhosphorImager (Molecular Dynamics, Sunnyvale, CA). After correcting for the background activity, the data were normalized and then the product concentration *versus* reaction times was plotted using the data analysis/graphics software KaleidaGraph™.

Preparation of oligonucleotides for melting studies

The template strand contains ϵ A, to which six complementary strands with different nucleotides opposite ϵ A were hybridized individually

Template	5'-GTCGACGXTCGCGAATTC-3'
Complement	3'-CAGCTGCYAGCGCTTAAG-5'

where X = ϵ A and Y = A, T, G, C, U.

To create an AP-site, the uracil containing complementary strand was modified. Uracil-DNA glycosylase was purchased from Gibco BRL, and used to remove the uracil from the 18-mer.⁵⁰⁻⁵¹ The 60 nmol of the U-strand was dissolved in buffer (1.16 ml, 50 mM Hepes-NaOH (pH 7.5), 0.1 mM EDTA), to which 40 units (40 μ l) of UDG was added, to a final strand concentration of 0.5 mM. After incubation of the mixture at 37 °C, for three hours, a reducing solution (133 μ l, 6 M NaBH₄, 0.2 M NaOH) was slowly added. After ten minutes, the reaction was quenched using 1 M HCl. The

product containing the reduced AP-site was then isolated using HPLC with a gradient from 0% to 25% (v/v) acetonitrile buffered with 0.1 M triethylamine ammonium acetate (TEAA), (pH 8.0). The flow rate was 1.0 ml/minute through a Waters Bondapak column (part no. WAT084176). The desired product, eluted at ~22 minutes was collected. The UDG mediated reaction on the uracil-template strand, after purification and dialysis, gave a yield of 5.4 A_{260} (~65%).

Each 18-mer was then desalted by dialysis for 48 hours, separated into 2.5 A_{260} aliquots and dried.

DNA melting

All melting studies were conducted using an AVIV 14DS UV-Vis spectrophotometer, which generates absorbance *versus* temperature profiles for up to five cells per melting experiment, as described.⁵² Secondary structure formation for each strand was predicted using DNA-MFOLD†. The A-complement strand was melted to validate the predictions generated by MFOLD.

Six melts were conducted with the duplex oligonucleotides containing ϵ A-strand against each complement including one containing AP-site. Dried samples were dissolved in buffer (100 μ l, 0.1 NaCl, 10 mM sodium cacodylate, 0.5 mM EDTA (pH 7.0)), and equimolar amounts of ϵ A template and complement strand were mixed.³³ Dilution of mixed strands was over an 80 to 100-fold scale (10^{-4} to 10^{-6} M), and the path length of the cuvettes varied between 0.1 cm to 0.8 cm. This allows for the concentration dependence of the T_m to be determined.

The absorbance at 260 nm for each dilution was recorded at 85 °C. From these values, and the mean extinction coefficient‡ of the two strands (ϵ A was assumed to have the same extinction coefficient as A and the extinction coefficient of AP was assumed to be zero), the concentration of each dilution was determined using Beer's law. The absorbance at 280 or 260 nm of each dilution was then recorded *versus* temperature, which was uniformly increased by 0.8 deg. C/minute, from 0 °C and 95 °C.

Profile analysis

The melting profiles generated were used to determine the thermodynamic parameters of duplex formation. Melting curves were fitted using the program MELTWIN v.3.0.³² MELTWIN applies the Van't Hoff analysis to determine thermodynamic parameters.³³ The thermodynamic parameters of duplex formation of normal mismatches in 18-mer oligonucleotides of the same sequences were computed using HYTHER™ server§.

Other methods

Proteins were quantified by Bradford reagent (BioRad) with bovine serum albumin as the standard. Oligonucleotides were synthesized with an Applied Biosystems model 394 DNA/RNA synthesizer. Adenine, thymine, guanine, and cytosine-CE phosphoramidites

were obtained from Applied Biosystems and special uracil and ethenoadenine-CE phosphoramidite from Glen Research.

Acknowledgments

We thank Dr S. H. Wilson for kindly providing purified UDG protein. We thank Norm Watkins for performing the UV melting experiment on the ϵ A-AP duplex. Supported in part by NIH R01 grants CA80917 (R.R.), HG02020 (J.S.L.), and CA53791, ES07572, CA31721 (S.M.), and by NIEHS Center grant ES06676.

References

- Friedberg, E. C., Walker, G. C. & Siede, W. (1995). *DNA Repair and Mutagenesis*, ASM Press, Washington, DC, USA.
- Krokan, H. E., Standal, R. & Slupphaug, G. (1997). DNA glycosylases in the base excision repair of DNA. *Biochem. J.* **325**, 1–16.
- Mitra, S., Hazra, T. K., Roy, R., Ikeda, S., Biswas, T., Lock, J. *et al.* (1997). Complexities of DNA base excision repair in mammalian cells. *Molecules Cells*, **7**, 305–312.
- Dizdaroglu, M., Karakaya, A., Jaruga, P., Slupphaug, G. & Krokan, H. E. (1996). Novel activities of human uracil DNA N-glycosylase for cytosine-derived products of oxidative DNA damage. *Nucl. Acids Res.* **24**, 418–422.
- Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E. & Tainer, J. A. (1995). Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell*, **80**, 869–878.
- Slupphaug, G., Eftedal, I., Kavli, B., Bharati, S., Helle, N. M., Haug, T. *et al.* (1995). Properties of a recombinant human uracil-DNA glycosylase from the UNG gene and evidence that UNG encodes the major uracil-DNA glycosylase. *Biochemistry*, **34**, 128–138.
- Stivers, J. T., Pankiewicz, K. W. & Watanabe, K. A. (1999). Kinetic mechanism of damage site recognition and uracil flipping by *Escherichia coli* uracil DNA glycosylase. *Biochemistry*, **38**, 952–963.
- Roy, R., Biswas, T., Hazra, T. K., Roy, G., Grabowski, D. T., Izumi, T. *et al.* (1998). Specific interaction of wild-type and truncated mouse N-methylpurine-DNA glycosylase with ethenoadenine-containing DNA. *Biochemistry*, **37**, 580–589.
- Ikeda, S., Biswas, T., Roy, R., Izumi, T., Boldogh, I., Kurosky, A. *et al.* (1998). Purification and characterization of human NTH1, a homolog of *Escherichia coli* endonuclease III. *J. Biol. Chem.* **273**, 21585–21593.
- Marenstein, D. R., Ocampo, M. T. A., Chan, M. K., Altamirano, A., Basu, A. K., Boorstein, R. J. *et al.* (2001). Stimulation of hNthl by YB-1 (DbpB): interaction between a base excision repair enzyme and a transcription factor. *J. Biol. Chem.* **276**, 21242–21249.
- Singer, B. & Hang, B. (1997). What structural features determine repair enzyme specificity and mechanism in chemically modified DNA? *Chem. Res. Tox.* **10**, 713–732.
- Roy, R., Brooks, C. & Mitra, S. (1994). Purification and biochemical characterization of recombinant N-methylpurine-DNA glycosylase of the mouse. *Biochemistry*, **33**, 15131–15140.

† Available on-line at

<http://mfold2.wustl.edu/~mfold/dna/form1.cgi>

‡ <http://paris.chem.yale.edu/extinct.html>

§ <http://jsl1.chem.wayne.edu/Hyther/Hytherm1.cgi>

13. Roy, R., Kennel, S. J. & Mitra, S. (1996). Distinct substrate preference of human and mouse *N*-methylpurine-DNA glycosylases. *Carcinogenesis*, **17**, 2177–2182.
14. O'Connor, T. R. (1993). Purification and characterization of human 3-methyladenine-DNA glycosylase. *Nucl. Acids Res.* **21**, 5561–5569.
15. Saparbaev, M. & Laval, J. (1994). Excision of hypoxanthine from DNA containing dIMP residues by the *Escherichia coli*, yeast, rat, and human alkylpurine DNA glycosylases. *Proc. Natl Acad. Sci. USA*, **91**, 5873–5877.
16. Dosanjh, M. K., Roy, R., Mitra, S. & Singer, B. (1994). 1,*N*⁶-ethenoadenine is preferred over 3-methyladenine as substrate by a cloned human *N*-methylpurine-DNA glycosylase (3-methyladenine-DNA glycosylase). *Biochemistry*, **33**, 1624–1628.
17. 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.
18. Lau, A. Y., Scharer, O. D., Samson, L., Verdine, G. L. & Ellenberger, T. (1998). Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell*, **95**, 249–258.
19. Lindahl, T., Sedgwick, B., Sekiguchi, M. & Nakabeppu, Y. (1988). Regulation and expression of the adaptive response to alkylating agents. *Annu. Rev. Biochem.* **57**, 133–157.
20. McCarthy, T. V., Karran, P. & Lindahl, T. (1984). Inducible repair of *O*-alkylated DNA pyrimidines in *Escherichia coli*. *EMBO J.* **3**, 545–550.
21. Bjelland, S., Birkeland, N. K., Benneche, T., Volden, G. & Seeberg, E. (1994). DNA glycosylase activities for thymine residues oxidized in the methyl group are functions of the AlkA enzyme in *Escherichia coli*. *J. Biol. Chem.* **269**, 30489–30495.
22. Bjoras, M., Klungland, A., Johansen, R. F. & Seeberg, E. (1995). Purification and properties of the alkylation repair DNA glycosylase encoded the MAG gene from *Saccharomyces cerevisiae*. *Biochemistry*, **34**, 4577–4582.
23. Mol, C. D., Arvai, A. S., Sanderson, R. J., Slupphaug, G., Kavli, B., Krokan, H. E. *et al.* (1995). Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell*, **82**, 701–708.
24. Savva, R., McAuley-Hecht, K., Brown, T. & Pearl, L. (1995). The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature*, **373**, 487–493.
25. 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.
26. Bruner, S. D., Norman, D. P. & Verdine, G. L. (2000). Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature*, **403**, 859–866.
27. Lau, A. Y., Wyatt, M. D., Glassner, B. J., Samson, L. D. & Ellenberger, T. (2000). Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase AAG. *Proc. Natl Acad. Sci. USA*, **97**, 13573–13578.
28. Verdine, G. L. & Bruner, S. D. (1997). How do DNA repair proteins locate damaged bases in the genome? *Chem. Biol.* **4**, 329–334.
29. Nyaga, S. G., Dodson, M. L. & Lloyd, R. S. (1997). Role of specific amino acid residues in T4 endonuclease V that alter nontarget DNA binding. *Biochemistry*, **36**, 4080–4088.
30. 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.
31. Abner, C. W., Lau, A. L., Ellenberger, T. & Bloom, L. B. (2001). Base excision and DNA binding activities of human alkyladenine DNA glycosylase are sensitive to the base paired with a lesion. *J. Biol. Chem.* **276**, 13379–13387.
32. McDowell, J. & Turner, D. H. (1996). Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: solution structure of (rGAGGUCUC)₂ by two-dimensional NMR and simulated annealing. *Biochemistry*, **35**, 14077–14089.
33. SantaLucia, J. (2000). The use of spectroscopic techniques in the study of DNA stability. In *Spectrophotometry and Spectrofluorimetry*, pp. 330–352, Oxford Press, Oxford.
34. Borer, P. N., Dengler, B., Tinoco, I., Jr & Uhlenbeck, O. C. (1974). Stability of ribonucleic acid double-stranded helices. *J. Mol. Biol.* **86**, 843–853.
35. Frier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T. & Turner, D. H. (1986). Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl Acad. Sci. USA*, **83**, 9373–9377.
36. Bommarito, S., Peyret, N. & SantaLucia, J. (2000). Thermodynamic parameters for DNA sequences with dangling ends. *Nucl. Acids Res.* **28**, 1929–1934.
37. Boulard, Y., Cognet, J. A. H. & Fazakerley, G. V. (1997). Solution structure as a function of pH of two central mismatches, C:T and C:C, in the 29 to 39 K-ras gene sequences, by magnetic resonance and molecular dynamics. *J. Mol. Biol.* **268**, 331–347.
38. Hang, B., Sagi, J. & Singer, B. (1998). Correlation between sequence-dependent glycosylase repair and the thermal stability of oligonucleotide duplexes containing 1,*N*⁶-ethenoadenine. *J. Biol. Chem.* **273**, 33406–33413.
39. Kouchakdjian, M., Eisenberg, M., Yarema, K., Basu, A., Essigmann, J. & Patel, D. J. (1991). NMR studies of the exocyclic 1,*N*⁶-ethenodeoxyadenosine adduct (ϵ dA) opposite thymidine in a DNA duplex. Non-planar alignment of epsilon dA(anti) and dT(anti) at the lesion site. *Biochemistry*, **30**, 1820–1828.
40. Leonard, G. A., McAuley-Hecht, K. E., Gibson, N. J., Brown, T., Watson, W. P. & Hunter, W. N. (1994). Guanine-1, *N*⁶-ethenoadenine base pairs in the crystal structure of d(CGCGAATT(ϵ dA)GCG). *Biochemistry*, **33**, 4755–4761.
41. Saparbaev, M., Kleibl, K. & Laval, J. (1995). *Escherichia coli*, *Saccharomyces cerevisiae*, rat and human 3-methyladenine DNA glycosylases repair 1,*N*⁶-ethenoadenine when present in DNA. *Nucl. Acids Res.* **23**, 3750–3755.
42. Barrett, T. E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J. & Pearl, L. H. (1998). Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. *Cell*, **92**, 117–129.
43. Verri, A., Mazzarello, P., Spadari, S. & Focher, F. (1992). Uracil-DNA glycosylases preferentially excise mispaired uracil. *Biochem. J.* **287**, 1007–1010.

44. Babic, I., Andrew, S. E. & Jirik, F. R. (1996). MutS interaction with mismatch and alkylated base containing DNA molecules detected by optical biosensor. *Mut. Res.* **372**, 87–96.
45. Biswas, I. & Hsieh, P. (1997). Interaction of MutS protein with the major and minor grooves of a heteroduplex DNA. *J. Biol. Chem.* **272**, 13355–13364.
46. Fazakerley, G. V., Quignard, E., Woisard, A., Guschlbauer, W., van der Marel, G. A., van Boom, J. H. *et al.* (1986). Structures of mismatched base pairs in DNA and their recognition by the *Escherichia coli* mismatch repair system. *EMBO J.* **5**, 3697–3703.
47. Marathias, V. M., Jerkovic, B. & Bolton, P. H. (1999). Damage increases the flexibility of duplex DNA. *Nucl. Acids Res.* **27**, 1854–1858.
48. Hess, M. T., Schwitter, U., Petretta, M., Giese, B. & Naegeli, H. (1997). Bipartite substrate discrimination by human nucleotide excision repair. *Proc. Natl Acad. Sci. USA*, **94**, 6664–6669.
49. Leblanc, B. & Moss, T. (1994). DNase I footprinting. In *Methods in Molecular Biology* (Kneale, G. G., ed.), vol. 30, pp. 1–10, Humana Press, Totowa, NJ.
50. Strauss, P., Beard, W., Patterson, T. & Wilson, S. (1997). Substrate binding by human apurinic/apyrimidinic endonuclease indicates a Briggs–Haldane mechanism. *J. Biol. Chem.* **272**, 1302–1307.
51. Fu, W., O’Handley, S., Cunningham, R. & Johnson, M. (1992). The role of the iron–sulfur cluster in *Escherichia coli* endonuclease III. A resonance Raman study. *J. Biol. Chem.* **267**, 16135–16137.
52. SantaLucia, J., Allawi, H. & Seneviratne, P. (1996). Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, **35**, 3555–3562.
53. SantaLucia, J. (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci. USA*, **95**, 1460–1465.
54. Fasman, G. (1975). *CRC Handbook of Biochemistry and Molecular Biology*, p. 597, CRC Press, Cleveland, OH.

Edited by K. Morikawa

(Received 15 November 2001; received in revised form 17 May 2002; accepted 17 May 2002)