Specific Interaction of Wild-Type and Truncated Mouse *N*-Methylpurine-DNA Glycosylase with Ethenoadenine-Containing DNA†

Rabindra Roy,* Tapan Biswas, Tapas K. Hazra, Gargi Roy, David T. Grabowski,‡ Tadahide Izumi, Ganesan Srinivasan, and Sankar Mitra

Sealy Center for Molecular Science and Department of Human Biological Chemistry and Genetics, The University of Texas Medical Branch, Galveston, Texas 77555

Received September 16, 1997; Revised Manuscript Received November 6, 1997

ABSTRACT: *N*-Methylpurine-DNA glycosylase (MPG), a ubiquitous DNA repair enzyme, is responsible for the removal of a wide variety of alkylated base lesions in DNA, e.g., *N*-alkylpurines and cyclic ethenoadducts of adenine, guanine, and cytosine. These lesions, some of which are mutagenic and toxic, are generated endogenously or by genotoxic agents such as *N*-alkylnitrosamines and vinyl chloride. Wild-type mouse MPG, expressed from recombinant baculovirus, was purified to near homogeneity for studying its specific interaction with substrate, 1,N^6-ethenoadenine- (ϵA-) containing DNA. Electrophoretic mobility shift assays (EMSA) indicated that MPG formed a specific complex with a 50-mer ϵA-containing duplex oligonucleotide. This complex was shown to be a transient reaction intermediate, because it could be formed only with the unreacted substrate and contained active enzyme molecules. DNA footprinting studies confirmed the specific binding of the protein to the ϵA-containing duplex oligonucleotide; eight nucleotides on the ϵA-containing strand and 16−17 nucleotides in the complementary strand spanning the base adduct were protected from DNase I digestion. A systematic deletion analysis of MPG was carried out in order to determine the minimally sized polypeptide capable of forming a stable substrate complex that is also suitable for characterization by NMR spectroscopy and X-ray crystallography. A truncated polypeptide (NΔ100CΔ18) lacking 100 and 18 amino acid residues from the amino and carboxyl termini, respectively, was found to be the minimal size that retained activity. The truncated and wild-type enzymes have similar kinetic properties. Moreover, both EMSA and DNase I footprinting studies indicated identical pattern of specific binding by the truncated and full-length polypeptides. Removal of five and nine additional residues from the amino- and carboxyl-termini of this polypeptide, respectively, resulted in a complete loss of activity. These results suggest that minimal structural change occured as a result of truncation in the NΔ100CΔ18 mutant, which may thus be suitable for elucidating the structure and mechanism of MPG.

Simple alkylating agents, such as methylmethanesulfonate and *N*-alkylnitrosamines, induce a variety of base alkyl adducts in DNA. The majority of these, including *N*-alkylpurines, e.g., 7-alkylguanine and 3-alkyladenine, are not mutagenic, although 3-methyladenine was shown to be toxic (*1*, *2*). After metabolic activation, vinyl chloride, an industrial byproduct and an environmental carcinogen, induces formation of cyclic ethenoadducts 1,N^6-ethenoadenine (ϵA), 1,N^2-ethenoguanine, N^2,3-ethenoguanine, and 3,N^4-ethenocytosine in DNA (*3*−*5*), some of which are mutagenic (*6*, *7*). A few of these adducts are also endogenously generated (*8*). All of these DNA adducts are repaired via the base excision repair pathway, whose first step is removal of the adduct by *N*-methylpurine−DNA glycosylase (MPG).[1] The resulting abasic (AP) site is then repaired in a series of steps starting with endonucleolytic cleavage of the AP site-containing DNA strand by AP-endonuclease (*9*). Although mammalian cells may possess multiple DNA glycosylases specific for alkylated base adducts (*10*), only MPG has so far been characterized, and it accounts for the major activity in removal of a variety of DNA base adducts (*11*). In particular, MPG is highly efficient in removing ϵA from DNA (*12*). Earlier efforts to purify mammalian MPGs to homogeneity from endogenous sources failed because of the gene's low expression at both mRNA (*13*) and protein levels (*14*). On the other hand, the fact that expression plasmids of human and rodent MPGs have been cloned by virtue of their ability to rescue MPG-

† This work was supported by USPHS Grants CA53791 and ES06676 (to S.M).

* To whom correspondence should be addressed: Sealy Center for Molecular Science, UTMB, Route 1079, Galveston, TX 77555-1079. Telephone (409) 772-1782; Fax (409) 747-8608; E-mail: rroy@scms.utmb.edu.

‡ Present address: Department of Cell and Structural Biology, The University of Texas Health Science Center, San Antonio, TX 78284.

[1] Abbreviations: A, adenine; DMS, dimethylsulfate; ϵA, 1,N^6-ethenoadenine; EMSA, electrophoretic mobility shift assay; hAPE, human apurinic/apyrimidinic (AP) endonuclease; IPTG, isopropyl β-D-thiogalactopyranoside; m^3A, 3-methyladenine; m^3G, 3-methylguanine; m^7G, 7-methylguanine; MPG, *N*-methylpurine−DNA glycosylase; PMSF, phenylmethanesulfonyl fluoride; T, thymine; UDG, uracil−DNA glycosylase; WT, wild-type; SDS−PAGE, sodium dodecyl sufate−polyacrylamide gel electrophoresis.

S0006-2960(97)02313-1 CCC: $15.00
Published on Web 01/13/1998

negative *Escherichia coli* indicates that recombinant MPGs are enzymatically active and could be used for biochemical studies (*15*, *16*). In our earlier effort we could not overproduce wild-type mouse MPG, as nonfusion polypeptides, at a high level in *E. coli*. As described in this report, we have now succeeded in expressing the full-length protein in insect cells and then purified it to near homogeneity.

An important step in understanding the mechanisms mediating MPG's recognition and repair of damaged lesions in DNA is to characterize the specific binding of the enzyme to damaged DNA. As described in this report, the availability of purified wild-type mouse MPG allowed us to investigate its interaction with an ϵA-containing duplex oligonucleotide substrate.

We have, recently, shown that MPG is organized in three distinct domains plus an unstructured region at the amino terminus (*17*). Because the unstructured region may hinder crystallization (*18*), we have determined the optimal size of the mouse MPG polypeptide that retains activity and specific binding. Size minimization and removal of unstructured regions should facilitate structural characterization by NMR spectroscopy and X-ray crystallography.

MATERIALS AND METHODS

Bacterial Strains and Plasmids. The prokaryotic expression vector pRSET (Invitrogen) and its host *E. coli* BL21-(DE3)[pLysS] (Novagen) were used for expression of truncated MPG polypeptides. The eukaryotic baculovirus expression vector pVL1393 (Invitrogen), linearized baculovirus DNA (BaculoGold, PharMingen, Inc.), and host Sf9 insect cells (Invitrogen) were used for expression of the full-length mouse MPG protein. The plasmid vector pET15b (Novagen) and the *E. coli* host BW528 (DE3) [xth^-nfo^-] were used for expression of human AP-endonuclease protein.

Construction of Expression Plasmids for Wild-Type and Truncated MPGs. The plasmid pGI1945 (*19*) was digested with *Eco*RI, and the cDNA fragment encoding full-length MPG was inserted into the *Eco*RI site of pVL1393 to generate recombinant plasmid pRmwBac for expression of wild-type (WT) MPG. The supercoiled plasmid pGI1945 containing mouse MPG cDNA (1062 bp; *19*) was used as the template for PCR-mediated construction of expression plasmids for truncated proteins. The DNA template (10 ng) was added to 100 μL of 1X PCR buffer [10 mM Tris-HCl (pH 8.3), 50 mM KCl, 2.3 mM $MgCl_2$, and 0.01% gelatin], containing 200 μM each of dCTP, dATP, dGTP, and dTTP and 50 pmol each of the two synthetic oligonucleotides. The sense primer was complementary to the 3′ end of the nontranscribed strand of mouse MPG cDNA and also included the recognition sequence for *Nde*I. The antisense primer was complementary to the 3′ end of the transcribed strand of the mouse MPG cDNA and also contained the recognition sequence for *Eco*RI. Each set of primers was designed to amplify a different length of the MPG cDNA fragment, encoding a protein that lacks different numbers of amino acid residues from both ends. The following oligodeoxynucleotides were used to generate the deletion mutants by PCR, with the amino- (N-) or carboxyl- (C-) terminal location and size of the expected deletions indicated: (i) CΔ9, 5′ CAA TTA AAA GGA ATT CCT AGC AGG CTG TTT GC 3′; (ii) CΔ18, 5′ CTG AGG GAA TTC CTA CTG TTC AGC CAC 3′ (iii) CΔ27, 5′ CTG AGG GAA TTC CTA GAC CCA TGG GCT 3′ (iv) NΔ96, 5′ CTC CGG CGG GAT CCC CAT ATG TCC AGC CCA GAG 3′; (v) NΔ100, 5′ GGA TCC TCA TAT GGA CCA TTC TGG C 3′; (vi) NΔ105, 5′ CCA GAG GAG GAT CCT CAT ATG CTA GGA CCA GAG 3′. The *Eco*RI and *Nde*I restriction sites are underlined in oligonucleotides i−iii and iv−vi, respectively. Taq DNA polymerase (2 units) from Boehringer Mannheim was added and DNA amplification was performed for 24 cycles (denaturation at 94 °C for 1 min, annealing at 55 °C for 1 min, and extension at 72 °C for 1 min for each cycle). After checking for the expected size on a 1% agarose gel, the PCR products were digested with *Nde*I and *Eco*RI and then purified using a PCR product purification kit (FMC Bioproducts). Finally, the products were subcloned to the *Nde*I and *Eco*RI sites of pRSETB in order to generate different expression plasmids, whose identities were confirmed by sequencing.

Expression and Assay of Recombinant Wild-Type and Truncated MPG Polypeptides. The pRmwBac expression plasmid for WT MPG and linearized baculovirus DNA were cotransfected into Sf9 cells by calcium phosphate precipitation (*20*). Transfer of mMPG cDNA sequences from pRmwBac to baculovirus DNA occurred by homologous recombination. Unlike the recombinant virus, the wild type cannot survive because it has a lethal deletion. Indications of viral infection were seen within 48 h, which included 25−50% increase in the cells' diameter and a marked increase in the size of the cell nuclei relative to cell size. At 96 h postinfection, the medium containing recombinant virus was collected by centrifugation and used to reinfect cells to increase the titer. The infected cells in the pellet was used for preparation of cell-free extract.

E. coli BL21(DE3)[pLysS] harboring MPG expression plasmids was grown at 37 °C to an A_{550} of 0.5−0.6 before addition of IPTG to a final concentration of 0.5 mM. The cells were allowed to grow for 16 h at 22 °C after addition of IPTG and then collected by centrifugation. The cell-free extracts were prepared by resuspending cells in TEDG buffer [20 mM Tris-HCl (pH 7.6), 1 mM EDTA, 0.5 mM DTT, and 5% glycerol] containing 150 mM NaCl and 0.2 mM PMSF, and followed by sonication on ice. After centrifugation at 4 °C for 15 min at 15000*g*, the supernatants were assayed for MPG activity by incubation with an excess of substrate, [*methyl*-^{3}H]-labeled methylated calf thymus DNA (370 cpm/μg of DNA; 109 fmol of m^7G and 20 fmol of m^3A in a total of 960 pmol of adenine and 640 pmol of guanine/μg of DNA), as described earlier (*21*). Because the insect cells and *E. coli* host contained endogenous MPG, the activity in Sf9 and BL21(DE3)pLysS cells harboring the empty vector was subtracted for each assay. One unit of MPG is defined as the activity needed to release 1 pmol of methylpurine/min at 37 °C under the assay conditions.

SDS−PAGE and Western Blot Analysis. Cell-free extracts (5−10 μg of soluble protein) of full-length and deletion mutants were subjected to SDS−PAGE (15% polyacrylamide) (*22*), and the resolved bands were stained with Coomassie brilliant blue. For Western blot analysis, proteins were transferred to a nitrocellulose membrane and the MPG bands were visualized with affinity-purified anti-mouse MPG polyclonal antibody (1:2000 dilution) using enhanced chemi-

luminescence (Amersham Life Sciences) detection technique according to the manufacturer's protocol.

Purification of WT and Optimal-Length MPG Polypeptides. Sf9 cells in a 400 mL suspension culture were infected with the recombinant virus carrying WT MPG at a multiplicity of infection of 5. The cells were harvested at 72 h postinfection. On the other hand *E. coli* BL21(DE3)[pLysS] harboring pRT118, the mouse NΔ100CΔ18 expression plasmid, was grown overnight at 37 °C in L-broth containing 100 μg/mL ampicillin and then diluted 50-fold with the same medium in 2 L and incubated at 37 °C until A_{550} reached about 0.6, when the culture was induced with IPTG (0.5 mM) for 10−12 h at 22 °C, and cells were then harvested.

The Sf9 and *E. coli* cell preparations were individually suspended in 40 mL of TEDG buffer containing 150 mM NaCl and 0.2 mM PMSF and then sonicated on ice. The lysates were centrifuged at 35000*g* for 30 min. One-tenth volume of Polymin P [5% (v/v), neutralized to pH 8.0] was added to the supernatant to precipitate nucleic acids, which were then removed by centrifugation at 12000*g* for 20 min. After addition of ammonium sulfate to 60% saturation and centrifugation 30 min later, the precipitates containing MPG were dissolved in 30−40 mL of TEDG buffer containing 0.1 M NaCl and dialyzed against 2 L of the same buffer at 4 °C with one change. The dialysate was then applied to a 25 mL S-Sepharose column. After extensive washing with TEDG buffer containing 0.1 M NaCl, the column was eluted with a NaCl gradient (0.1−1.0 M) in TEDG buffer; MPG-containing fractions eluted at 0.25−0.3 M NaCl were pooled, dialyzed against TEDG buffer, and then applied to a MonoS HR 5/5 column (Pharmacia FPLC system) equilibrated with TEDG buffer. After washing, the column was developed with a linear gradient (0−0.5 M) of NaCl (flow rate 1.5 mL/min). The MPG was eluted at 0.45 M NaCl as the major peak. Fractions containing MPG activity were pooled and stored in aliquots at −80 °C. On average, 5−6 mg of mouse wild-type MPG and 10−12 mg of minimal-length MPG were recovered from 400 mL of recombinant baculovirus-infected insect (Sf9) cells and 1 L of *E. coli* cells, respectively.

Purification of Human AP Endonuclease. The hAPE cDNA was introduced into the pET15b plasmid vector (Novagen) in its *Nde*I and *Xho*I sites to generate the expression plasmid. The construct carried an extra 20 amino acid residues at the N-terminus, including a histidine hexamer to enable purification through a nickel column. *E. coli* BW528 (DE3) harboring the expression plasmid was grown in a fermenter at 37 °C until the A_{600} reached 1.5, when the culture was induced with 0.5 mM IPTG at 25 °C for 4 h. Cells were harvested, resuspended in TN buffer [20 mM Tris-HCl (pH 8.0) and 300 mM NaCl] and lysed with a French press. The extract was applied on a 10 mL TALON column (Clontech), washed with 100 mL of TN buffer, and eluted with TN buffer containing 10 mM imidazole. The eluents containing hAPE were then applied on 3 mL of Ni-NTA resin (Qiagen), washed with 30 mL of TN buffer containing 40 mM imidazole, and eluted with 6 mL of TN buffer containing 200 mM imidazole. The eluted pooled fractions containing hAPE were concentrated to about 6−10 mg/mL by Centricon-10 (Amicon) and applied on a Superdex 75 (Pharmacia) gel-filtration column equilibrated with a buffer containing 20 mM Tris-HCl (pH 8.0), 300 mM NaCl, 5% glycerol, and 1 mM DTT. Fractions containing hAPE were pooled and stored in aliquots at −80 °C.

Electrophoretic Mobility Shift Assay. Oligonucleotides (50-mer) containing a single εA or A residue at position 27 and the complementary sequence containing a T opposite εA or A were synthesized using standard phosphoramidite chemistry. The εA-containing and standard oligonucleotides were deprotected in 14 N ammonium hydroxide for 16 h at room temperature and 55 °C, respectively. The sequence of the oligonucleotide was 5′ TCG AGG ATC CTG AGC TCG AGT CGA CG(εA) TCG CGA ATT CTG CGG ATC CAA GC 3′. The single-stranded oligonucleotides were purified on a denaturing 10% polyacrylamide gel before 5′-end-labeling with T4 polynucleotide kinase and [γ-^{32}P]ATP. Complementary strands were then annealed in TEN buffer [10 mM Tris-HCl (pH 7.8), 1 mM EDTA, and 100 mM NaCl; *18*] and the duplex oligonucleotide was purified through a G-50 column to remove the unincorporated nucleotides. An AP-site-containing oligonucleotide was generated by treating uracil- (U-) containing 50-mer duplex oligonucleotide with an excess amount of uracil−DNA glycosylase (U.S. Biochemical Corp.). The U-oligonucleotide had identical sequence as the εA-oligonucleotide except for substitution of εA with U. The binding reaction was carried out in an assay buffer containing 25 mM Hepes-KOH (pH 7.8), 0.5 mM EDTA, 0.5 mM DTT, 150 mM KCl, 10% glycerol, 1 μg of poly(dI·dC), 4−5 pmol of pure protein, and 5−10 fmol of the ^{32}P-labeled oligonucleotide in a final volume of 20 μL. Unlabeled oligonucleotides of the same sequence (20 pmol) were added as competitors when necessary. After 10−15 min at room temperature, 5 μL of the reaction mixture was electrophoresed on a 6% nondenaturing polyacrylamide gel in TBE buffer (90 mM Tris-borate, pH 8.0, and 2 mM EDTA) at 10 V/cm for 150 min at 4−5 °C. The gel was transferred to DE81 paper, dried, and exposed to X-ray film (Kodak, Biomax).

Enzymatic Activity of MPG Present in MPG−Substrate Complex. The binding reaction using unlabeled εA-containing oligonucleotide (40 pmol) and NΔ100CΔ18 MPG (20 pmol) was performed as described above, and the whole reaction mixture was electrophoresed on a 6% nondenaturing polyacrylamide gel in 1 × TAE buffer [40 mM Tris-acetate (pH 8.3) and 1 mM EDTA]. The shifted band complex was identified using a ^{32}P-labeled oligonucleotide as the marker in a separate lane in the same gel. The protein was eluted from the gel using a Gel-Spin protein recovery unit (Worthington Biochemical Corp.). The protein-containing eluate was adjusted to the buffer for MPG assay (as described earlier) in 200 μL and incubated at 37 °C with ^{32}P-labeled εA-oligonucleotide (∼30 000 cpm) along with a control in a reaction mixture complete except for the protein. Aliquots (48 μL) were withdrawn at different times, heated at 68 °C for 5 min, and subsequently incubated with 0.5 μg of AP-endonuclease at 37 °C for 20 min after the concentration of Mg^{2+} was adjusted to 5 mM. The control was incubated for 16 h at 37 °C and then treated as for the other samples. After the reaction mixture was extracted twice with phenol/chloroform, the radiolabeled oligonucleotide was precipitated with 2.5 volumes of chilled ethanol in the presence of 0.3 M sodium acetate, pH 7.0, and 2 μg of yeast tRNA. The precipitate was washed with 70% ethanol, dried, dissolved in loading buffer containing 66% formamide and 0.03 N

NaOH, and electrophoresed on a sequencing gel containing 8% polyacrylamide/7 M urea.

Kinetics of MPG Reaction with ϵA-Containing Duplex Oligonucleotide Substrate. WT (16.7 nM) and NΔ100CΔ18 (13.4 nM) MPGs were individually incubated with ϵA-containing duplex oligonucleotide substrate (50-mer, 35 nM−4.6 μM) for 5 min at 37 °C in MPG assay buffer (as described above) in a total volume of 20 μL. The reaction was stopped by inactivating the enzyme at 68 °C for 5 min. The products containing the AP sites were quantitatively cleaved into smaller fragments by incubating with a large excess amount (0.5 μg) of AP-endonuclease at 37 °C for 20 min after the concentration of Mg^{2+} was adjusted to 5 mM. The reaction mixture was then deproteinized, ethanol-precipitated, and resolved on a 8% polyacrylamide sequencing gel as described above. The radioactivity in the bands of the gel was subsequently quantified with a PhosphorImager (Molecular Dynamics, Sunnyvale, CA). The data were fitted to the Michaelis−Menten equation by a nonlinear least-squares method using the data analysis/graphics application program KaleidaGraph, version 3.0, for the Macintosh.

DNase I Footprinting Reactions. DNase I footprinting was performed essentially as described by Leblanc and Moss (*23*). The binding reaction (50 μL) was performed with 4 fmol of ^{32}P-end-labeled oligonucleotide (∼10 000 cpm), 0.5 μg of poly(dI·dC), and varying amounts of WT or NΔ100 CΔ18 MPG in a buffer containing 20 mM Tris-HCl (pH 8.0), 1 mM EDTA, 1 mM DTT, 60 mM KCl, and 10% glycerol. After incubation for 10 min at room temperature, the DNA was digested with 5 μL of 0.002 unit (Kunitz)/μL DNase I (Sigma) for 2 min. The reaction was stopped by addition of reaction stop buffer (1% SDS, 200 mM NaCl, 20 mM EDTA, pH 8.0, and 40 μg/mL tRNA), and after extraction with phenol/chloroform, the DNA was precipitated in ethanol and then resolved by electrophoresis on a sequencing gel.

Determination of Binding Affinities. A major DNase I band in the complementary strand, protected by WT and NΔ100 CΔ18 MPGs, was selected for quantification of radioactivity in the PhosphorImager. The protein concentration that reduced the radioactive counts in the band of interest to 50% was considered as the K_D.

Other Methods. Proteins were quantitated by the bicinchoninic acid procedure (*24*) with bovine serum albumin as the standard. Radioactivity in liquid samples was quantitated in a Beckman Model LS6000SC liquid scintillation counter, using Scintiverse II (Fisher Scientific) as scintillant. Oligodeoxynucleotides were synthesized in an Applied Biosystems Model 394 DNA/RNA synthesizer. Adenine, thymine, guanine, and cytosine-CE phosphoramidites were obtained from Applied Biosystems and special ethenoadenine-CE phosphoramidite from Glen Research.

RESULTS

Expression and Purification of WT MPG. Because we could not overexpress the full-length mouse MPG as a nonfusion protein in the bacterial system, we used insect cells (Sf9) and a baculovirus expression vector (pVL1393) system to overexpress the protein. The protein was purified and shown by SDS−PAGE to be nearly homogeneous (Figure 1, lane WT mMPG).

Specific, Transient Complex between MPG and Substrate Oligonucleotide. We showed earlier that, when incubated at 22 °C with a 25-mer duplex oligonucleotide containing a single ϵA residue, the human MPG in cell-free extract is capable of forming a specific complex, which was detectable as a slower migrating band in EMSA (*12*). Figure 2A shows that the catalytically active full-length MPG also binds to a 50-mer ϵA-containing oligonucleotide (lane 2). This binding was specific, since it was effectively competed with a 1000-fold molar excess of unlabeled ϵA-oligonucleotide (lane 3) but not with control adenine-containing oligonucleotide of identical sequence (lane 4).

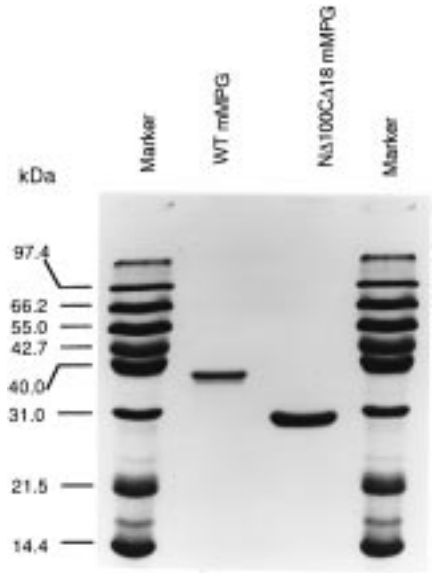

FIGURE 1: SDS−PAGE (15% polyacrylamide) analysis of purified full-length and NΔ100CΔ18 MPGs. Markers in the outside lanes are mid-range molecular weight marker proteins (14 400−97 400).

Because MPG does not require any cofactor, its continued reaction could not be prevented during EMSA. Therefore, it was important to prove that the observed complex was indeed the transient reaction intermediate. We utilized several distinct approaches to confirm the identity of the complex. First, we observed that the complex formation at 37 °C (Figure 2B, lane 2) was significantly reduced (as determined by the relative amount of radioactivity in the shifted complex) compared to that at 22 °C. This difference suggested that the steady-state level of the enzyme−substrate complex was reduced when the substrate was converted into the product and that the abasic site product did not form a specific complex with the enzyme. We then incubated ϵA-oligonucleotide (100 fmol) with an excess amount of MPG (20 pmol) at 37 °C for 30 min and purified the oligonucleotide by phenol/chloroform extraction and alcohol precipitation. An aliquot of the oligonucleotide was then treated with hot piperidine (*10*); it was found that MPG removed most of the ϵA residues to generate AP sites that were cleaved by the alkaline treatment (data not shown). This oligonucleotide did not give a gel-shifted band after incubation with MPG at 22 °C (Figure 2B, lane1). It should be noted that the presence of a few higher bands (Figure 2B, lanes 1 and 2) other than MPG−substrate complex could be due to trace contamination of sequence-specific proteins (which are not detectable on the Coomassie-stained gel) in the WT MPG preparation. We then showed directly that the AP-oligonucleotide did not form an EMSA-detectable complex with MPG. An oligonucleotide of sequence identical to the ϵA-

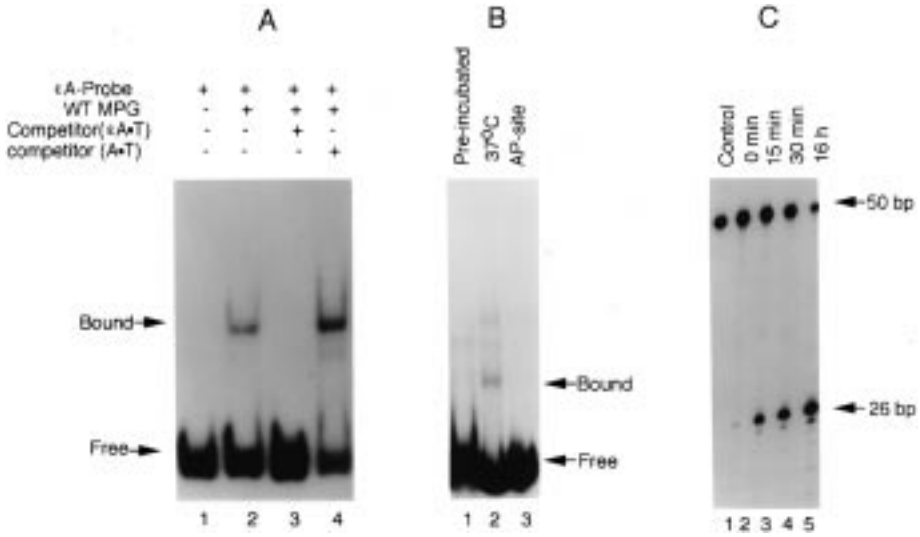

FIGURE 2: EMSA of WT MPG with εA-containing 50-mer duplex oligonucleotide. (A) WT MPG was incubated with ^{32}P-labeled εA-oligonucleotide and then electrophoresed as described in Materials and Methods. Unlabeled duplex oligonucleotides of identical sequences but containing either εA or A were used as competitors and are shown in parentheses. (B) The ^{32}P-labeled εA-oligonucleotide was incubated with the WT MPG under different conditions as follows: (Preincubated) ^{32}P-Labeled εA-oligonucleotide was preincubated with excess MPG at 37 °C for 20 min, extracted with phenol/chloroform and then chloroform, dried, dissolved in water, and used for binding with MPG as described in panel A. (37 °C) Binding was performed as in panel A but at 37 °C. (AP-site) Binding was performed as in panel A but with a 50-mer oligonucleotide containing an AP site instead of εA. The AP-oligonucleotide was prepared by treating U-oligonucleotide with an excess amount of uracil-DNA glycosylase and then treated as described under preincubated. The positions of free DNA probe (free) and enzyme-bound DNA complexes (bound) are indicated. (C) Enzymatic activity of MPG eluted from MPG−substrate complex. An εA-containing 50-mer duplex oligonucleotide was incubated in the absence (lane 1) or presence (lanes 2−5) of eluted MPG for various time periods (0−16 h). 50 bp, substrate; 26 bp, product.

oligonucleotide (except for the lack of a base at the εA site) was incubated with MPG. The absence of a significant complex (Figure 2B, lane 3) shows that MPG forms a detectable complex only with the substrate and not with the reaction product.

Finally, we ruled out the possibility that the shifted complex in EMSA was formed with inactive MPG. After electrophoresis of the reaction mixture of MPG and unlabeled εA-oligonucleotide duplex, we cut out the complex band in polyacrylamide and extracted the protein from the gel. The protein was then incubated with fresh 5′-^{32}P-labeled εA-containing oligonucleotide and the reaction kinetics were monitored by electrophoresis of DNA after treatment with an excess amount of AP-endonuclease that cleaves the abasic site in the oligonucleotide generated by MPG (Figure 2C). Time-dependent increase in the amount of 26-mer oligonucleotide product indicated the presence of active MPG in the first shifted complex. Thus the transient steady-state intermediate of the MPG reaction could be identified by EMSA.

Footprinting Analysis of MPG-Bound Substrate. The MPG substrate complex was further characterized by DNA footprinting analysis. The DNase I footprint for WT MPG bound to εA-containing oligonucleotide is shown in Figure 3.

A footprint of eight nucleotides was observed spanning the εA residue, with 4 nucleotides on the 5′ side and four nucleotides on the 3′ side of εA in the lesion-containing strand (Figure 3A, lanes 9−11). The presence of a 26-mer cleaved product (5′ PO_4 and 3′OH) among the footprinted bands (Figure 3A, lanes 10 and 11, indicated by an arrow) of the εA-containing strand strongly suggests the binding of active protein to DNA because it was likely to result from alkali-induced hydrolysis of the abasic site generated by the glycosylase. The complementary strand showed a larger footprint spanning about 16−17 nucleotides (Figure 3A, lanes 6−8). Protection from DNase I was distinct and definitive on both strands, but some nucleotides were protected to a much lower extent than others. Such hypersensitive sites were observed in both the εA-containing and complementary strands. No protection (footprinting on the complementary strand is only shown in Figure 3A, lanes 3−5) was observed for the unmodified oligonucleotide, a result that provides further evidence for the specific interaction of MPG with the substrate. Figure 3B shows a schematic representation of the protected region in the 50-mer oligonucleotide. In separate experiments we performed footprinting with a 45-mer oligonucleotide having a completely different sequence and containing εA at the 23rd position. The observed pattern of protection was similar to that with the 50-mer oligonucleotide (data not shown), which indicated that the protection was damage-specific and did not depend on the nucleotide sequence.

Analysis of Deletion Mutants of MPG. A series of deletion constructs (Figure 4A) was made in the coding sequence of the mouse MPG cDNA, and the truncated proteins were expressed in *E. coli* BL21(DE3)[pLysS]. The glycosylase activity of truncated MPGs expressed in *E. coli* BL21(DE3)-[pLysS] from different deletion constructs was measured in cell-free extracts and corrected for background activity (Figure 4A). Deletion of 100 and 18 amino acid residues from the amino and carboxyl termini, respectively, caused no inactivation of the enzyme. However, further deletion of five amino acid residues from the amino terminus not only led to complete loss of solubility but the insoluble protein was apparently degraded to a smaller size [Figure 4B, lane NΔ105CΔ9 (P)]. The insoluble aggregate was denatured and renatured according to the method of Sam-

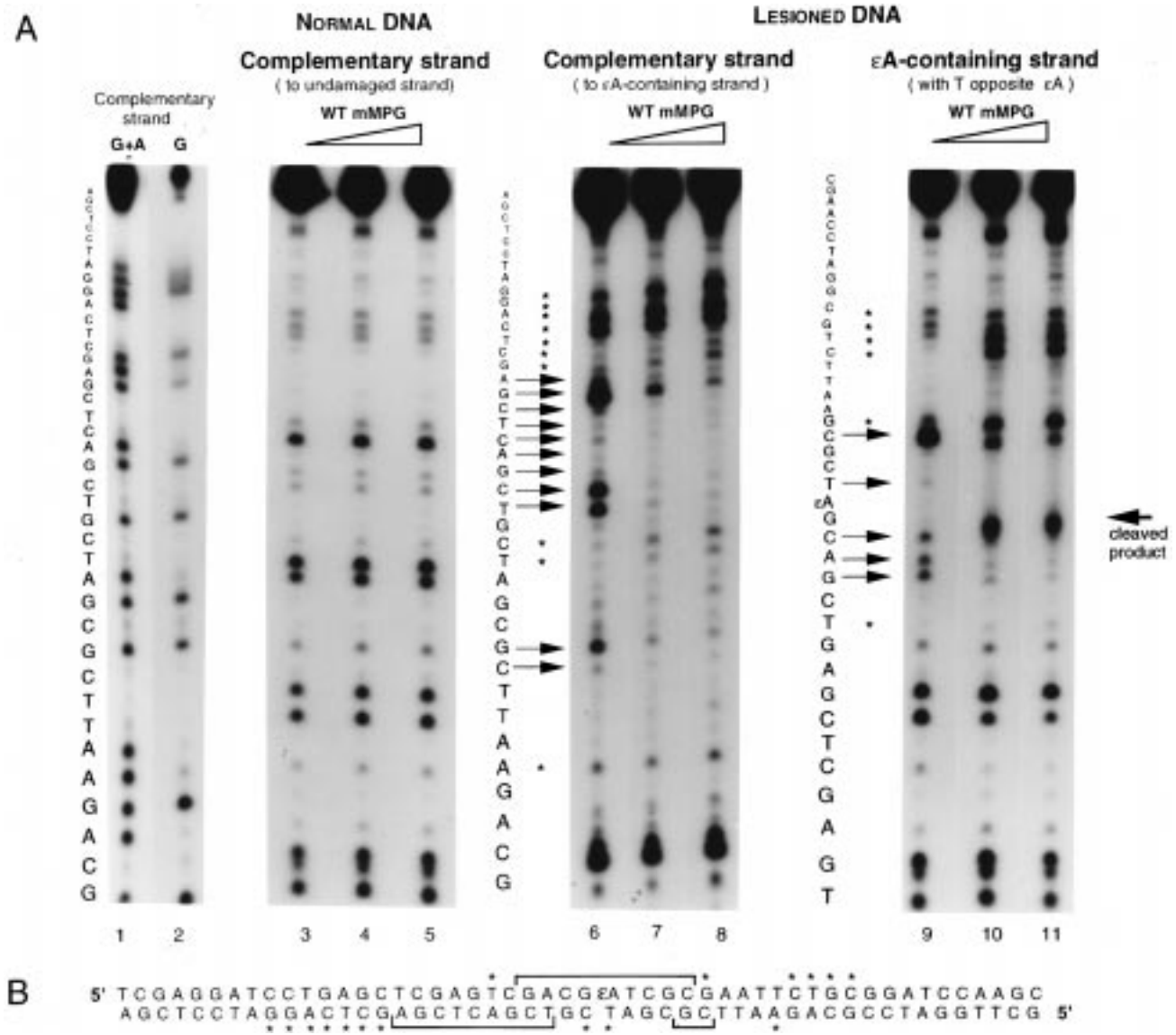

FIGURE 3: DNase I footprinting of an ϵA-containing oligonucleotide with WT MPG. An ϵA-containing 50-mer oligonucleotide was incubated in the absence or presence of WT MPG and subjected to DNase I footprinting as described in the experimental procedures. (A) Lanes 1 and 2 represent G and G + A Maxam−Gilbert sequencing reactions, respectively, on the complementary strand. Lanes 3−5 show the DNase I digestion pattern of the complementary strand in the normal DNA (containing no damaged base) in the presence of 0, 7, and 14 pmol of WT MPG. Lanes 6−8 and 9−11 show the protection pattern in the complementary or ϵA-containing strand, respectively, with 0, 7, and 14 pmol of WT MPG. Protected bands are shown with arrows, and DNase I hypersensitive sites are marked with asterisks. The cleaved product in the ϵA-containing strand is indicated with a bold arrow. (B) Protected regions in both strands are bracketed and DNase I hypersensitive sites are marked with asterisks.

brook et al. (*25*). The protein became soluble as evident from Western blotting (data not shown); however, the MPG activity was below the level of detection. As a control, the insoluble aggregate of NΔ100CΔ18 MPG, similarly purified from the pellet fraction of sonicated *E. coli*, was subjected to the same denaturation/renaturation procedure. The protein became soluble as judged by SDS−PAGE followed by Coomassie staining and Western blotting; the solubilized protein showed detectable activity (data not shown). This result supports our conclusion that NΔ105CΔ9 MPG could be inactive even when present in the soluble form. Deletion of nine additional residues from the carboxyl terminus also led to reduced solubility of the protein [Figure 4B, lane NΔ96CΔ27 (S)], and the soluble fraction failed to show any detectable activity (Figure 4A). NΔ100CΔ18 is thus the optimal polypeptide in terms of solubility and enzymatic activity.

Kinetic Properties of WT and Optimal-Length Proteins. The optimal-size protein NΔ100CΔ18 could be purified to apparent homogeneity in large amounts needed for future structural studies, e.g., by X-ray crystallography and NMR spectroscopy (Figure 1, lane NΔ100CΔ18 mMPG), and the kinetic properties of WT and NΔ100CΔ18 MPGs were compared. Both the proteins had comparable specific activity (200 units/mg for full-length and 260 units/mg for NΔ100CΔ18 MPG) while utilizing methylated bases as substrates. These enzymes also released methylated bases and ϵA in a dose- and time-dependent manner. The active enzyme concentration (16.7 nM for full-length and 13.4 nM for NΔ100CΔ18 mutant) and time (5 min) of incubation (data not shown) chosen for K_m and k_{cat} determination were in the linear range of enzyme activity. The K_m, k_{cat}, and the catalytic efficiency (k_{cat}/K_m) for both full-length and NΔ100CΔ18 polypeptides were also comparable for the

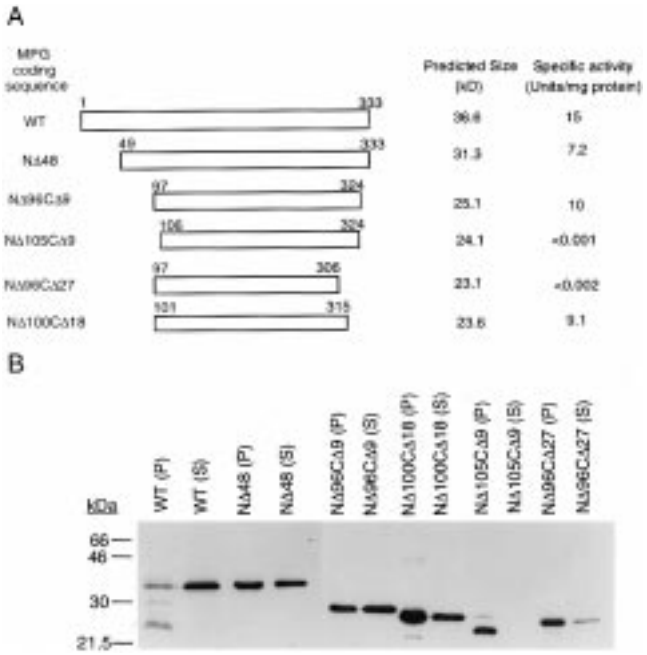

FIGURE 4: (A) Deletion constructs of mouse MPG. The numbers at amino (N) and carboxyl (C) termini correspond to the first and last amino acid residues in the truncated polypeptides. A baculovirus expression plasmid was used to produce the WT protein as described in Materials and Methods. The rest of the constructs are pRSET-based, and *E. coli* BL21(DE3)[pLysS] was used for protein expression. Specific activities were measured in cell-free extracts, and the predicted sizes of mutant proteins were calculated from their coding sequences. (B) Western blot analysis of WT and deletion mutants of MPG. Blots of insect cell Sf9 and *E. coli* extracts containing full-length and mutant mouse MPGs were probed with anti-MPG antibody as described in Materials and Methods. The marker lane included high molecular weight prestained rainbow proteins (14 300−200 000) from Amersham.

Table 1: Kinetic Parameters of Recombinant Mouse MPGs for the Binding to and Excision of ϵA from ϵA-Containing Duplex Oligonucleotide Substrate[a]

enzyme	K_m (nM)	k_{cat} (min^{-1})	k_{cat}/K_m (min^{-1} nM^{-1})	K_D (nM)
wild type	40−100	∼1	0.025−0.01	1
NΔ100CΔ18	60−130	∼0.6	0.01−0.004	0.2−0.4

[a] Experimental Details are described in Materials and Methods.

substrate ϵA-containing oligonucleotide (Table 1). The mouse MPG appears to be catalytically more efficient in releasing ϵA in the present study than m^{3}A, computed previously (*21*), and shows consistency with our earlier observation (*12*). Although m^{3}A was present at multiple sites along with other methylated bases in calf thymus DNA (*21*), in the present study ϵA was present as a single modified base in an oligonucleotide. The K_D values were determined for both wild-type and NΔ100CΔ18 mutant MPG from substrate (ϵA-containing oligonucleotide) binding followed by DNase I footprinting, and the mutant protein had 2−5-fold higher affinity than WT MPG (Table 1). The K_D value appears to be higher than that of human MPG (*26*).

Specific Binding of Truncated Mutant (NΔ100CΔ18) to ϵA-Containing Oligonucleotide Substrate. The truncated mutant (NΔ100CΔ18) of MPG behaved nearly identically to the WT protein in EMSA. It was bound specifically to a 50-mer ϵA-containing oligonucleotide because the binding was effectively competed with a 1000-fold molar excess of unlabeled ϵA-containing oligonucleotide, but not with the control A-oligonucleotide (Figure 5A). The truncated protein also showed similar protection of substrate DNA from DNase as the WT MPG (Figure 5B, lanes 4−6). Because the optimally truncated active mouse MPG protein (∼23 kDa) showed no significant loss of excision activity, and large amounts can be produced in pure form, we conclude that this protein, in both free and DNA-bound form, may be ideal for structural studies.

DISCUSSION

The ubiquitous nature of MPG reflects the importance of repairing the specific base lesions that are removed by this enzyme. However, the promiscuous activity of this repair protein in recognizing purine base lesions of widely different structures is surprising. Because MPG appears to be the predominant if not the only enzyme responsible for removing multiple mutagenic and carcinogenic cyclic ethenoadducts of adenine, guanine, and cytosine in DNA (*27*), the study of substrate recognition by the mammalian enzyme is of particular importance. *E. coli* AlkA is also active toward many of the same lesions as MPG, but the enzymes' substrate ranges are not identical (*28*).

The MPGs' biochemical mechanisms have not yet been elucidated. In a previous study we used EMSA to demonstrate the specific binding of an ϵA-containing oligonucleotide substrate to partially purified mammalian MPG (*12*). In this report we show that the binding of both purified WT and minimal-length MPG to DNA is damage- (ϵA-) specific, because the binding could not be competed with an A-containing oligonucleotide having the same sequence as the substrate. Surprisingly, the presence of unlabeled nonsubstrate competitor increased the binding of MPG with the labeled substrate DNA (Figure 2A, lane 4). We routinely observe this not only in the case of MPG but also for other DNA repair proteins, e.g., O^6-methylguanine−DNA methyltransferase (*29*). The explanation could be that adsorption of the trace amount of radiolabeled substrate, present in the reaction mixture, to the microfuge tube wall is reduced by an excess of unlabeled DNA, leading to an increase of the specific binding of MPG to the substrate. Moreover, the complex appears to be a transient intermediate formed before the cleavage of the glycosyl bond, because preincubating of the substrate with MPG or performing the binding reaction at 37 °C instead of 22 °C reduced the extent of binding. Furthermore, no complex was observed between the enzyme and an AP site, and the reaction product and the complex were also shown to contain active enzyme molecules. The complex of MPG with ϵA-oligonuleotide could be trapped and detected in EMSA presumably because of the slow reaction rate at 22 °C. One of the major limitations of analyzing the interaction between DNA repair enzymes and their substrate DNA is the lack of substrate analogues that are recognized but not acted upon by the enzyme. In the case of *E. coli* endonuclease III (Nth protein) and MutM proteins, which are DNA glycosylases for oxidized bases, noncleavable substrates were used to produce substrate−enzyme dead-end complexes (*30*, *31*). These enzymes and T4 endonuclease V (endoV) have, in addition to DNA glycosylase activity, AP lyase activity, due to which they form transient covalent intermediates via Schiff bases. Stable

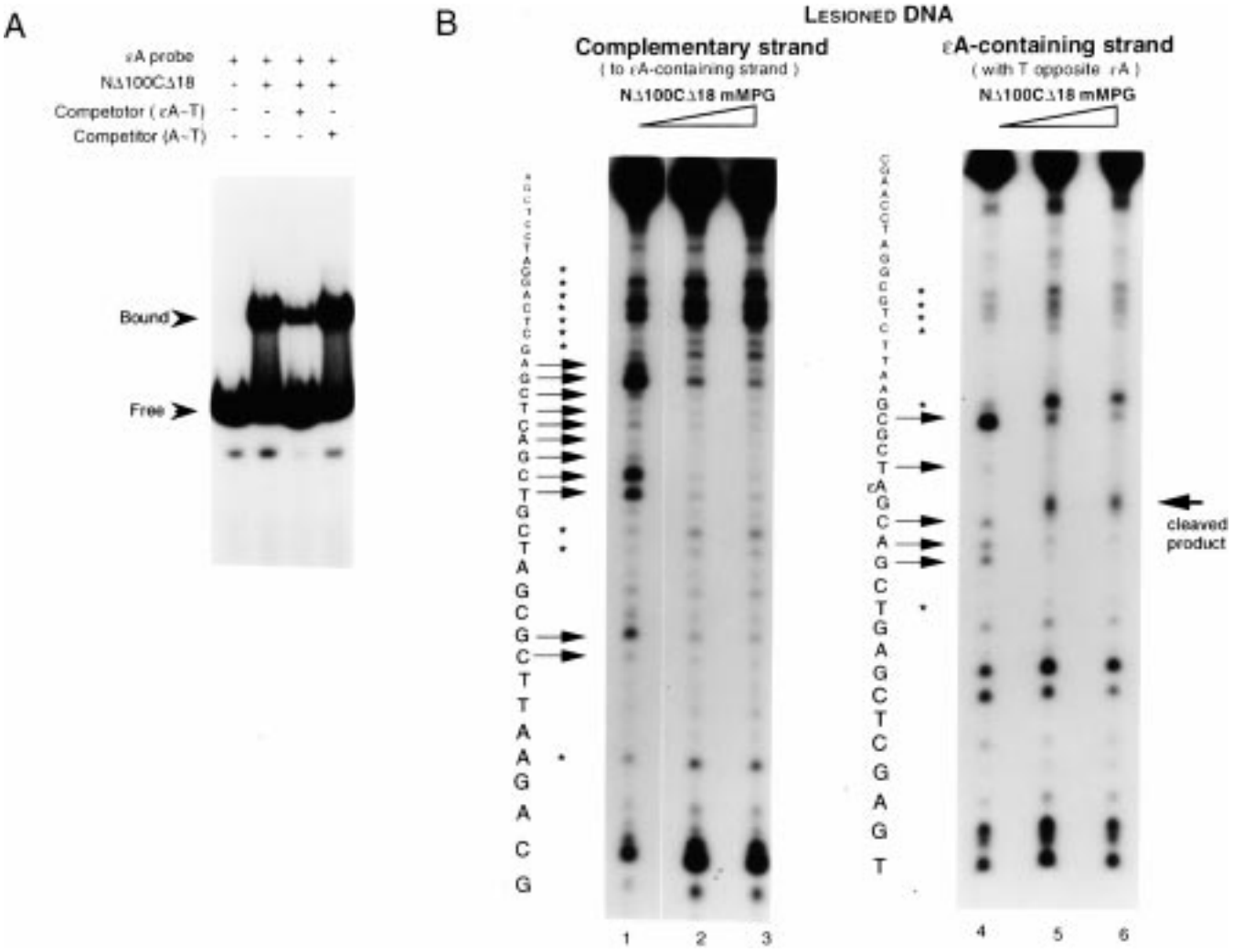

FIGURE 5: (A) EMSA of NΔ100CΔ18 MPG with ϵA-containing 50-mer duplex oligonucleotide. The experimental procedure is described in Materials and Methods and other details are same as in Figure 2A. (B) DNase I footprinting of the ϵA-containing 50-mer with NΔ100CΔ18 MPG. An ϵA-containing 50-mer was incubated in the presence of 0, 10, and 20 pmol of truncated MPG and subjected to DNase I footprinting as described in Materials and Methods. Lanes 1−3 and 4−6 show the protection pattern in the complementary strand or ϵA-containing strand in the absence or presence of NΔ100CΔ18 MPG, respectively.

substrate enzyme complexes can be isolated by reducing the Schiff base intermediates (*31*, *32*). Simple DNA glycosylases, such as MPG without AP lyase activity, do not form Schiff bases as reaction intermediates (*33*). In this study we show the presence of a transient reaction intermediate of a simple DNA glycosylase. We recently observed electrophoretic gel mobility shift of MGMT, a DNA repair protein that acts stoichiometrically, when incubated with its substrate, O^6-methylguanine-containing duplex oligonucleotide (*29*). We propose that similar EMSA and DNA footprint analysis can be performed for other DNA repair enzymes whose turnover numbers are either intrinsically low or can be reduced by experimental manipulations.

The recognition mechanisms of sequence-specific DNA-binding regulatory proteins have been elucidated in many cases (*34*, *35*). However, little is known about the mechanisms by which DNA repair proteins recognize and bind to their cognate DNA lesions. The DNase I footprinting analysis showing that MPG protects a short stretch (∼8 base pairs on the ϵA strand) suggests that there are probably few contact sites between MPG and the damaged DNA. This result is similar to those obtained for other DNA repair enzymes, e.g., *E. coli* endonuclease III (*30*), FPG (*31*), and endoV (*32*). In all cases so far, the footprints of regulatory proteins are usually at least twice as long those observed with the repair proteins on both strands. The smaller binding site with fewer DNA−protein contacts probably reduces the possibility of sequence dependence, since a single modified base, regardless of the neighboring sequence, should be the determinant for recognition of DNA repair proteins.

To our knowledge, substrate recognition of only two DNA repair proteins, namely, uracil−DNA glycosylase (UDG) and endoV, has been investigated in detail by X-ray crystallography (*36*, *37*). However, both of these proteins have stringent substrate specificity. Because MPG has a broad substrate range, its mechanism of recognition may be different from that of UDG or endoV.

Elucidation of the tertiary structure of MPG and the MPG−substrate DNA complex by NMR and X-ray crystallography will be essential for a comprehensive understanding of the mechanism of this enzyme. In particular, because NMR analysis is currently feasible only with polypeptides smaller than 20 kDa, we have determined the approximate boundaries of nondeactivating truncation in order to identify the smallest active protein for NMR studies. The minimal length polypeptide, NΔ100CΔ18, is nearly identical to the

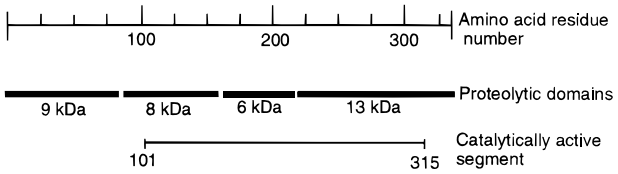

FIGURE 6: Linear functional map of MPG. The proteolytic domains and the catalytically active segment are aligned with the WT protein.

full-length protein in enzymatic activity as judged by their comparable K_m and k_{cat} values.

Surprisingly, MPG polypeptides shorter than the NΔ100-CΔ18 polypeptide, especially truncated at the amino terminus, form insoluble aggregates when expressed in *E. coli* and are also susceptible to proteolytic degradation. It is likely that the tertiary structures of these shorter polypeptides were seriously perturbed.

We have recently shown that the mouse MPG is organized in several protease-resistant domains, including an unstructured region (5 kDa) at the amino terminus, linked sequentially to three domains with molecular masses of 8, 6, and 13 kDa, respectively (*17*). All three domains appear to be necessary for substrate recognition and catalysis. The present deletion mutagenesis study provides further insight into the structural organization of the protein. The NΔ100CΔ18 polypeptide lacks the unstructured 5 kDa region and part of the 8 kDa domain. The alignment of the smallest active polypeptide, NΔ100CΔ18, with the proteolytic domains has been schematically represented in Figure 6.

It appears likely that the catalytic residues are located in regions spanning amino acid residues 101−105 and 307−315. During catalysis, the cleavage of an *N*-glycosidic bond may involve formation of an intermediate ester that would require Asp, Glu, or Cys as the catalytic residue (*15*). The three-dimensional structure of AlkA has recently been solved, and Asp-238 was proposed to be the catalytic residue (*38*, *39*). It is intriguing that aspartic acid is present in both the pentapeptide 101−105 (DHSGR) and the carboxyl-terminal nonapeptide 307−315 (SVVDRVAEQ) of mouse MPG. Site-directed mutagenesis might confirm whether these residues are indeed involved in ϵA recognition and catalysis. Furthermore, identification of site-specific MPG mutants with altered substrate specificity should help elucidate the mechanism of the MPG reaction.

Finally, the identical specificity of the optimal length polypeptide, NΔ100 CΔ18, and the WT MPG and preferred substrate binding of the truncated protein make this mutant an excellent candidate for studying the structure and reaction mechanism of MPG.

ACKNOWLEDGMENT

We thank Dr. Thomas Wood and Mr. Chinappa D. Kodira at the Recombinant DNA Laboratory (RDL) in the Sealy Center for Molecular Science for synthesizing the oligonucleotides and for DNA sequence analysis. We also thank Dr. David Konkel for critically reading the manuscript and Ms. Wanda Smith for excellent secretarial help.

REFERENCES

1. Boiteux, S., Huisman, O., and Laval, J. (1985) *EMBO J. 25*, 2569−2573.
2. Larson, K., Sharn, J., Shenkar, R. and Strauss, B. (1985) *Mutat. Res. 233*, 211−218.
3. Laib, R. J. (1986) in *The Role of Cyclic Nucleic Acid Adducts in Carcinogenesis and Mutagenesis* (Singer, B., & Bartsch, G., Eds.) IARC Scientific Publication *70*, pp 101−108, Oxford University Press, Oxford, England.
4. Leonard, N. J. (1984) *CRC Crit. Rev. Biochem. 15*, 125−199.
5. Leonard, N. J. (1992) *Chemtracts: Biochem. Mol. Biol. 3*, 273−297.
6. Singer, B., Kusmierek, J. T., Folkman, W., Chavez, F., and Dosanjh, M. K. (1991) *Carcinogenesis 12*, 745−747.
7. Cheng, K. C., Preston, B. D., Cahill, D. S., Dosanjh, M. K., Singer, B., and Loeb, L. A. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9974−9978.
8. Marnett, L. J., and Burcham, P. C. (1993) *Chem. Res. Toxicol. 6*, 771−785.
9. Friedberg, F. C., Walker, G. C., and Siede, W. (1995) Base excision repair, in DNA Repair and Mutagenesis, pp 135−176, ASM Press, Washington, DC.
10. Hang, B., Chenna, A., Rao, S., and Singer, B. (1996) *Carcinogenesis 17*, 155−157.
11. Mitra, S., Hazra, T. K., Roy, R., Ikeda, S., Biswas, T., Lock, J., Boldogh, I., and Izumi, T. (1997) *Mol. Cells 7*, 305−312.
12. Dosanjh, M. K., Roy, R., Mitra, S., and Singer, B. (1994) *Biochemistry 33*, 1624−1628.
13. Roy, G., Roy, R., and Mitra, S. (1997) *Anal. Biochem. 246*, 45−51.
14. Male, R., Haukanes, B. I., Helland, D. E., and Kleppe, K. (1987) *Eur. J. Biochem. 165*, 13−19.
15. Chakravarti, D., Ibeanu, G. C., Tano, K., and Mitra, S. (1991) *J. Biol. Chem. 266*, 15710−15715.
16. Samson, L., Derfler, B., Boosalis, M., and Call, K. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9127−9131.
17. Roy, R., Kumar, A., Lee, J. C., and Mitra, S. (1996) *J. Biol. Chem. 271*, 23690−23697.
18. Joachimiak, A., and Sigler, P. B. (1991) *Methods Enzymol. 208*, 82−99.
19. Tatsuka, M., Ibeanu, G., Izumi, T., Narayan, S., Ramana, C. V., Kim, N. K., Kang, W., Roy, G., and Mitra, S. (1995) *DNA Cell Biol. 14*, 37−45.
20. Summers, M. D., and Smith, G. R. (1987) *A manual of methods for the baculovirus vectors for insect cell culture procedures*, Texas Agricultural Experiment Station, Bulletin 1555, Texas A&M University, College Station, TX.
21. Roy, R., Brooks, C., and Mitra, S. (1994) *Biochemistry 33*, 15131−15140.
22. Laemmli, U. K. (1970) *Nature 227*, 680−685.
23. Leblanc, B., and Moss, T. (1994) *Methods Mol. Biol. 30*, 1−10.
24. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Porvenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. (1985) *Anal. Biochem. 150*, 76−85.
25. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: a laboratory manual*, pp 646−648, *Cold Spring Harbor Laboratory, Cold Spring Harbor, NY*.
26. Scharer, O. D., and Verdine, G. L. (1995) *J. Am. Chem. Soc. 117*, 10781−10782.
27. Dosanjh, M. K., Chenna, A., Kim, E., Frankel-Conrat, H., Samson, L., and Singer, B. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 1024−1028.
28. Helland, D. E., Male, R., Haukanes, B. I., Olsen, L., Haugan, I., and Kleppe, K. (1987) *J. Cell Sci.* (*Suppl.*) *6*, 139−146.
29. Hazra, T. K., Roy, R., Biswas, T., Grabowski, D. T., Pegg, A. E., and Mitra, S. (1997) *Biochemistry 36*, 5769−5776.
30. O'Handley, S., Scholes, C. P., and Cunningham, R. P. (1995) *Biochemistry 34*, 2528−2536.
31. Tchou, J., Michaels, M., Miller, J. H., and Grollman, A. P. (1993) *J. Biol. Chem. 268*, 26738−26744.
32. Latham, K. A., Taylor, J.-S., and Lloyd, R. S. (1995) *J. Biol. Chem. 270*, 3765−3771.
33. Dodson, M. L., Michaels, M. L., and Lloyd, R. S. (1994) *J. Biol. Chem. 269*, 32709−32712.
34. Seeman, N. C., Rosenberg, J. M., and Rich, A. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 804−808.

35. Steitz, T. A. (1990) *Q. Rev. Biophys. 23*, 205−280.
36. Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E., and Tainer, J. A. (1996) *Nature 384*, 87−92.
37. Vassylyev, D. G., Kashiwagi, T., Mikami, Y., Ariyoshi, M., Iwai, S., Ohtsuka, E., and Morikawa, K. (1995) *Cell 83*, 773−782.
38. Yamagata, Y., Kato, M., Odawara, K., Tokuno, Y., Nakashima, Y., Matsushima, N., Yasumura, K., Tomita, K.-I., Ihara, K., Fujii, Y., Nakabeppu, Y., Sekiguchi, M., and Fujii, S. (1996) *Cell 86*, 311−319.
39. Labahn, J., Scharer, O. D., Long, A., Ezaz-Nikpay, K., Verdine, G. L., and Ellengerger, T. E. (1996) *Cell 86*, 321−329.

BI972313L