Specific Recognition of O^6-Methylguanine in DNA by Active Site Mutants of Human O^6-Methylguanine-DNA Methyltransferase[†]

Tapas K. Hazra,[‡] Rabindra Roy,[‡] Tapan Biswas,[‡] David T. Grabowski,[‡,§] Anthony E. Pegg,[∥] and Sankar Mitra*[,‡]

Sealy Center for Molecular Science and Department of Human Biological Chemistry and Genetics, University of Texas Medical Branch, Route 1079, Galveston, Texas 77555, and Department of Cellular and Molecular Physiology and Pharmacology, Pennsylvania State University College of Medicine, P.O. Box 850, Hershey, Pennsylvania 17033

Received December 16, 1996; Revised Manuscript Received March 10, 1997[⊗]

ABSTRACT: O^6-Methylguanine-DNA methyltransferase (MGMT), a ubiquitous DNA repair protein, acts as a monomer in removing the mutagenic DNA adduct O^6-alkylguanine (induced by alkylating carcinogens) via a stoichiometric reaction. The alkyl group is transferred without a cofactor to a specific cysteine acceptor residue of MGMT, Cys-145 in the case of human MGMT, containing 207 amino acid residues and thereby inactivates the protein. As a prelude to the investigation of the reaction mechanism of human MGMT by elucidation of its structure in free and substrate-bound forms via NMR spectroscopy and X-ray crystallography, two types of MGMT mutants were generated and characterized. First, systematic deletion analysis of the protein was carried out to determine the smallest size at which it is active or inactive but forms a stable complex with the substrate and so may be useful for NMR spetroscopic analysis. Deletion of more than 8 or 31 residues from the amino or carboxyl terminus, respectively, led to the loss of both activity and substrate binding. Removal of Arg-9 or Leu-176 and distal residues inactivated the protein, presumably by altering its tertiary structure. On the basis of the criteria of bacterial overexpression and solubility, the mutant MGMT with deletion of 28 residues at the carboxyl terminus should be suitable for NMR studies. In the second approach, we examined mutants at the active site (Cys-145) that retain substrate binding. Inactive C145A and C145S substitution mutants were found to form specific and stable complexes with an O^6-methylguanine (m^6G)-containing oligonucleotide substrate. Wild type MGMT also formed a similar complex, but only as a transient intermediate. Footprinting studies indicated a strong discriminatory effect of the base adduct on the binding of C145A to substrate DNA; 17−18 nucleotides on the m^6G-containing strand and 13−14 nucleotides in the complementary strand spanning the base adduct were protected from DNase I digestion by the mutant protein. These results, as well as the identical protease sensitivity of the wild type and mutant proteins, suggest minimal structural change due to conservative mutations at the active site. Thus, the mutant proteins may be utilized for solving the structure and mechanism of human MGMT.

Cellular DNA is subject to attack by a variety of environmental alkylating agents, and to endogenous methylation resulting in a number of alkylated base adducts (Singer & Grunberger, 1983; Loveless, 1969). Although O^6-alkylguanine is a relatively minor adduct among these lesions, it is generally accepted as the major mutagenic and carcinogenic adduct because of its preference for pairing with thymine over cytosine during DNA replication. Thus, this lesion may lead to GC → AT mutation (Loveless, 1969; Snow et al., 1984; Loechler et al., 1984). A ubiquitous and unusual repair protein O^6-methylguanine-DNA methyltransferase (MGMT,[1] EC 2.1.1.63), repairs O^6-alkylguanine via an unusual mechanism in that it accepts the alkyl group on a unique cysteine residue in a single-step stoichiometric reaction, and is thereby inactivated (Foote et al., 1980; Olsson & Lindahl, 1980). MGMT acts as a monomer and does not require any cofactor in restoring the original guanine without DNA synthesis (Demple et al., 1982; Boulden et al., 1987; Bhattacharyya et al., 1990). MGMT plays a critical role in cytotoxicity of antitumor alkylating agents that are derivatives of (2-chloroethyl)-*N*-nitrosourea (CNU; Brent et al., 1985). Because MGMT removes the primary DNA lesion induced by these drugs, cellular drug resistance is often dependent on the MGMT level, which varies widely in different normal as well as tumor cells (Mitra & Kaina, 1993).

The major *Escherichia coli* MGMT (the product of the *ada* gene, thus also called the Ada protein) was the first MGMT to be discovered (Foote et al., 1980; Olsson & Lindahl, 1980). The 39 kDa protein, with two domains of nearly identical size, also acts as a transcriptional activator of the genes in the *ada* regulon. While the N-terminal domain is involved in transcriptional regulation, the 19 kDa C-terminal domain possesses MGMT activity (Lindahl et al.,

[†] This research was supported by U.S. Public Health Service Grants CA31721, ES07572, and ES06676 (to S.M.) and CA18137 (to A.E.P.).

* To whom correspondence should be addressed: University of Texas Medical Branch, Sealy Center for Molecular Science, Route 1079, Galveston, TX 77555.

[‡] University of Texas Medical Branch.

[§] Present address: Department of Cellular and Structural Biology, University of Texas Health Science Center at San Antonio, San Antonio, TX 78284.

[∥] Pennsylvania State University College of Medicine.

[⊗] Abstract published in *Advance ACS Abstracts,* May 1, 1997.

[1] Abbreviations: CNU, (2-chloroethyl)-*N*-nitrosourea; EMSA, electrophoretic mobility shift assay; FPLC, fast protein liquid chromatography; IPTG, isopropyl β-thiogalactopyranoside; MGMT, O^6-methylguanine-DNA methyltransferase; hMGMT, human MGMT; m^6G, O^6-methylguanine; NMR, nuclear magnetic resonance; PCR, polymerase chain reaction; WT, wild type.

S0006-2960(96)03085-1 CCC: $14.00

1988). *E. coli* has a second, constitutive MGMT, the 19 kDa Ogt, expressed at a low level (Potter et al., 1987; Rebeck et al., 1988).

Several eukaryotic MGMT cDNAs have been cloned and characterized (Tano et al., 1990; Hayakawa et al., 1990). These proteins are similar in size to Ogt and the C-terminal fragment of the Ada protein. A comparison of the amino acid sequences of various eukaryotic and bacterial MGMTs shows that the methyltransferases do not have extensive sequence homology except for the universal presence of a pentapeptide sequence, PCHRV, which includes the alkyl acceptor cysteine residue, and some conserved residues scattered throughout the proteins (Moore et al., 1994; Morgan et al., 1993).

The structure of the 19 kDa C-terminal half of the Ada protein was recently elucidated by X-ray crystallography (Moore et al., 1994). Although this structure has provided some insight into the mechanism of MGMT, detailed information about substrate recognition and the alkyl transfer reaction is still lacking, particularly for mammalian MGMTs. The mammalian proteins are significantly different from the bacterial protein in their ability to recognize O^6-benzylguanine as a substrate (Dolan et al., 1991).

Although it reacts inefficiently with free O^6-alkylguanine base, MGMT's biological substrate is the base adduct in DNA (Dolan et al., 1991). In order to recognize O^6-alkylguanine in DNA, the protein needs to scan DNA by binding to it reversibly. The mechanisms of this binding and translocation are not understood. Chan et al. (1993) calculated that the difference in the affinity of human MGMT (hMGMT) for substrate vs control DNA was about 3-fold. Such a small discrimination makes it difficult to elucidate the mechanism of MGMT from physicochemical studies of the protein−DNA complex.

We decided to isolate hMGMT mutants that retain specific substrate recognition and are suitable for biophysical studies, using two approaches. First, we carried out a systematic investigation of active truncation mutants of hMGMT and also inactive mutants that may still retain specific substrate binding. Size reduction has a significant advantage in the elucidation of structures of the free protein and protein−DNA complex by NMR spectroscopy (Wüthrich, 1996). Earlier studies showed that removal of 10 and 36 residues from the amino and carboxyl terminus of hMGMT inactivated the protein but that 28 residues at the carboxyl terminus were dispensable for activity (Crone et al., 1994; Morgan et al., 1993). However, correlations of size with activity and substrate binding were not studied. In the second approach, we decided to examine hMGMT−DNA interaction by identifying inactive mutants that retain specific substrate recognition and can be investigated by X-ray crystallography and NMR spectroscopy. Because the alkyl acceptor Cys-145 is the active site residue in the 207-amino acid hMGMT, we examined active site mutants for their ability to form specific and stable complexes with the substrate DNA.

Here we report that in truncation mutants activity and substrate binding could not be separated, although an 19 kDa active, truncated protein may be suitable for NMR studies. However, the active site mutants C145A and C145S form stable complexes with the substrate and would be suitable for structural studies of protein−DNA interaction.

MATERIALS AND METHODS

The prokaryotic expression vectors pRSET (Invitrogen) and *E. coli* BL21 (DE3) pLysS (Novagen) lysogen were used for the expression of full-length and truncated human MGMT polypeptides.

Construction of Expression Plasmids. The pKT100 plasmid containing the complete coding sequence for 207 amino acid residues of hMGMT (Tano et al., 1990) was used as the template for PCR-mediated construction of expression plasmids for the truncated proteins. Wild type and deleted coding sequences were amplified with 5−10 ng of pKT100 DNA template (linearized with *Eco*RI) in a 100 μL PCR mixture containing 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 2.5 mM $MgCl_2$, 0.01% gelatin, four dNTPs (each at 200 μM), and 50 pmol of two oligonucleotide primers. The primer sequences corresponded to distinct hMGMT coding sequences. The sense primer contained an *Nde*I linker sequence at the 5′ terminus, and the antisense primer contained the *Bam*HI linker, along with an ochre (TGA) nonsense codon sequence. The sequences of the oligonucleotides are shown below; the mutant proteins are named to denote the number of amino acid residues deleted at the amino (N) or the carboxyl (C) end. For example, NΔ7 denotes the mutant with deletion of 7 amino acid residues at the N terminus. sense full length, 5′GTA CGA ATT CAT ATG GAC AGG GAT3′; antisense full length, 5′TAC ATG GAT CCT TAT GCA CAT ACT CAG TT3′; NΔ7, 5′CAA GGG AAT TCA TAT GAA ACG CAC C3′; NΔ8, 5′GGA TCC TCA TAT GCG CAC CAC ACT G3′; NΔ9, 5′GGA TCC TCA TAT GAC CAC ACT GGA C3′; CΔ28, 5′GCT CGG ATC CTA TCA TGG CTT CCC3′; CΔ31, 5′CAA CGG GAT CCT TCA CAA CCG GTG3′; CΔ32, 5′GCC TGG ATC CTA TCA CCG GTG GCC3′. *Nde*I and *Bam*HI sites, underlined in the sense and antisense oligonucleotides, respectively, were used for PCR cloning of full-length and truncated proteins. Amplification was carried out with 2 units of Taq DNA polymerase (Boehringer) for 24 cycles using a program of 1.5 min at 94 °C for denaturation, 1 min at 55 °C for annealing, and 1 min at 70 °C for DNA replication. After the predicted size of the PCR product was confirmed by electrophoresis in 1% agarose, the DNA was purified using the "Spin-Bind" PCR purification system (FMC). Amplified DNAs were then digested with *Nde*I and *Bam*HI, subjected to a second cycle of purification, and then ligated to the pRSETB expression vector linearized with *Nde*I and *Bam*HI. The identity of the expected coding sequences was confirmed by DNA sequencing and physical mapping. This plasmid construction strategy allowed expression of wild type and truncated recombinant hMGMT proteins as nonfusion products.

Construction of Site-Specific Mutants of hMGMT. The mutants R9A, L176A, C145S, and C145A were generated using Stratagene's Chameleon Double-Stranded Site-Directed Mutagenesis Kit. This kit is based on a modification of the unique site elimination mutagenesis procedure described by Deng and Nickoloff (1992). Here we altered the unique *Sca*I site in the *amp* gene of the vector. The primer used to mutate the *Sca*I site was 5′CTG GTG AGT ATT CAA CCA AGT3′. The underlined base was changed so that the codon GAG was changed to GAA without affecting its specificity. The other mutagenic primers were 5′GTC CAG TGT GGT GGC TTT CAT TTC ACA3′ for R9A, 5′GAC CAC TCT

GTG GCT CGG GAT GAG GAT3′ for C145S, 5′GAC CAC TCT GTG GGC CGG GAT GAG GAT3′ for C145A, and 5′GCC TGG CTT CCC AGC CCG GTG GCC TTC3′ for L176A, where the mismatches are underlined. R9A and L176A mutations were created both in full-length and in minimal-length (NΔ8CΔ31) constructs; when NΔ8CΔ31 was used as the template, the mutagenic primer was 5′GTC CAG TGT GGT GGC CAT ATG TAT ATC3′ for R9A and 5′GAG CTC GGA TCC ATC CCG GTG GCC TTC3′ for L176A. The sequence corresponding to the entire hMGMT amino acid coding region was checked by DNA sequencing to ensure that only the desired mutations were present.

Expression and Assay of Recombinant Wild Type and Mutant hMGMT Polypeptides. *E. coli* BL21 (DE3) pLysS-containing hMGMT expression plasmids were grown to an A_{550} of 0.4−0.5 before adding IPTG to 0.5 mM, except to cultures with NΔ8CΔ31 and NΔ9CΔ31 expression plasmids. We decided to induce the latter mutants with 50 μM IPTG, which have very low expression in *E. coli*, at a slower rate than the other mutants on the basis of prior optimization studies. The cells were allowed to grow for 5−6 h at room temperature after addition of IPTG and then collected by centrifugation. Extracts were prepared from the cells after resuspension at 0 °C in buffer A [20 mM Tris-HCl (pH 7.5), 1 mM DTT, 1 mM EDTA, and 5% glycerol] containing 100 mM NaCl, followed by sonication. Cell debris was pelleted by centrifugation at 4 °C for 15 min at 15 000*g*, and the supernatants were assayed for hMGMT activity by measuring the transfer of [*methyl*-^{3}H]label from calf thymus DNA (methylated by reaction with [^{3}H]-*N*-methyl-*N*-nitrosourea) to hMGMT itself. The methylated protein was digested with proteinase K, and radioactivity was recovered in the alcohol-soluble fraction (Boulden et al., 1987; Waldstein et al., 1982). Because of *E. coli* host used was MGMT-positive, the activity in BL21 (DE3) pLysS cells containing the empty vector was subtracted for each assay. One unit of methyltransferase was defined as the amount that causes the removal of 1 pmol of methyl groups from O^6-methylguanine in DNA.

SDS−PAGE and Western Blot Analysis. Four micrograms of protein in the total soluble extract of *E. coli* containing each deletion mutant was resolved by 15% sodium dodecyl sulfate−polyacrylamide gel electrophoresis (Laemli, 1970) and transferred to a nitrocellulose membrane (Schleicher & Schuell). The hMGMT band in the blot was visualized with anti-hMGMT polyclonal antibody raised in rabbits (1:400 dilution) by the ECL technique (Amersham), used according to the manufacturer's protocol.

Purification of hMGMT Polypeptides and Mutant Proteins. Several colonies of freshly transformed *E. coli* BL21 (DE3) pLysS with expression plasmids for WT or different mutant proteins were each inoculated into 2 L of L-broth containing 200 μg/mL ampicillin and grown with vigorous shaking at 37 °C to an A_{550} of 0.7−0.8 before IPTG was added. The cultures were continued overnight (14−15 h) at 22−23 °C. The cells were harvested by centrifugation, and hMGMT was purified from sonicated extracts in a series of steps starting with the removal of nucleic acids with polymin P as previously described (Roy et al., 1995; Shiota et al., 1992; von Wronski et al., 1991), except that the final DNA cellulose affinity chromatography step was replaced by FPLC on a MonoS column, resulting in a nearly homogeneous preparation. Approximately 10 mg of protein was loaded onto a MonoS column equilibrated with buffer A containing 50 mM NaCl. The average yield (except C145S) of nearly homogeneous proteins was 7−8 mg per 5−6 g of packed cells; the yield of C145S was 2−3 mg from the same amount of cell. On the basis of the stoichiometry of methyl acceptor activity, 60−70% of purified MGMT molecules were active.

Electrophoretic Mobility Shift Assay (EMSA). A 27-mer oligonucleotide containing a single O^6-methylguanine at the 17th position was purchased from National Biosciences; the sequence of the modified oligonucleotide was 5′CCA GTG AAT TCC CGG G m^{6}GA TCC GTA GAC3′. The complementary strand contained cytosine opposite the guanine adduct. An oligonucleotide with the identical sequence except for substitution of O^6-methylguanine by guanine was used as a control for testing nonspecific binding in EMSA.

The oligonucleotides were labeled with ^{32}P at the 5′ termini using [γ-^{32}P]ATP and polynucleotide kinase and then annealed to the complementary strand in a buffer containing 77 mM Tris-HCl (pH 7.8) and 2.3 mM $MgCl_2$. The specific and nonspecific DNA−protein complexes between the oligonucleotide and different hMGMT proteins were analyzed by electrophoretic mobility shift assay (EMSA; Dosanjh et al., 1994). The specific binding reaction was carried out in a buffer (20 μL) containing 25 mM Hepes (pH 7.8), 0.5 mM EDTA, 0.5 mM DTT, 50 mM NaCl, and 10% glycerol along with 8−10 fmol of labeled oligonucleotide, 2.5 pmol of unlabeled oligonucleotide, 1 μg of poly(dI·dC), and 100 pmol of purified full-length protein (WT, C145S, or C145A) or partially purified inactive truncated mutants NΔ7CΔ32 (150−200 pmol) and NΔ9CΔ31 (125−150 pmol). Quantitation of these proteins was based on comparative band intensity in Western blotting, with homogeneous WT hMGMT being used as the standard. Unlabeled duplex competitor oligonucleotides (25 pmol) were added, when necessary, before the labeled oligonucleotide, which was added last to the reaction mixture. The association constants were determined by measuring the free and bound probe as described earlier (Stone et al., 1991).

To examine further the specificity of complex formation, the supershift assay with WT protein was carried out in the presence of anti-hMGMT polyclonal antibody. After 10 min of incubation of WT protein with the labeled probe, 2 μL of antiserum was added and the sample was incubated further for 10 min at 20 °C. For control experiments, nonimmune rabbit serum was used instead of immune serum. The antibody was also tested for any binding to the labeled probe under the same condition.

Nonspecific binding was carried out with three different purified hMGMT (WT, C145S, and C145A) proteins without poly(dI·dC) in a 15 μL reaction volume using a control oligo of the same sequence, except for substitution of O^6-methylguanine with guanine. Each reaction mixture contained 220 pmol of protein. After incubation for 10 min at 37 or 20 °C, as indicated, a 10 μL aliquot for specific binding and 7.5 μL for the nonspecific binding reaction were electrophoresed in TBE buffer [0.09 M Tris-borate (pH 8.0) and 2 mM EDTA] at 150 V. The gel was prerun at constant voltage (150 V) for 40 min before the samples were applied. After completion of electrophoresis, the gel was fixed in 10% methanol and 10% acetic acid, dried, and then exposed to X-ray film (Kodak, Biomax). The amount of radioactivity in the unbound and bound oligonucleotide was quantified by exposing the gel to a phosphorImager.

Controlled Proteolysis. Digestion of 5 μg of wild type hMGMT C145A and C145S mutant proteins was carried out with 0.05 μg of trypsin in a buffer (15 μL) containing 25 mM Tris-HCl (pH 8.0) and 0.1 M NaCl at 22 °C. Aliquots of the digested products were analyzed at various times by SDS−PAGE in 20% polyacrylamide, followed by staining with Coomassie brilliant blue.

DNase I Footprinting Reactions. A 45-mer oligonucleotide containing a single O^6-methylguanine at the 22nd position was purchased from National Biosciences, Inc. The sequence of the oligonucleotide was 5′GCT CCC TCT GAA GGC TCC AGG m^{6}GAA GAG TGT CCT CTG CTC CCT CCG3′. The complementary strand contains cytosine opposite the guanine adduct. An oligonucleotide with the identical sequence except for substitution of O^6-methylguanine by guanine was used as a control strand for testing any nonspecific protection during footprinting.

DNase I footprinting was performed essentially as described by Leblanc and Moss (1994). The oligonucleotide containing m^{6}G was 5′-end-labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP and was purified on a 10% polyacrylamide gel. An equimolar amount of complementary strand was annealed to produce a duplex oligonucleotide for DNase I digestion. The binding reaction (25 μL) was performed with 1−2 fmol of ^{32}P-end-labeled oligonucleotide (5000−6000 cpm), 1.7 pmol of cold oligo, 1 μg of poly(dI·dC), and 25−50 pmol of C145A mutant protein in a buffer containing 20 mM Tris-HCl (pH 8.0), 1 mM EDTA, 1 mM DTT, 100 mM NaCl, and 10% glycerol. After incubation for 10 min at 37 °C, the DNA was digested with 5 μL of 0.01 ku/μL DNase I (Sigma) for 2 min before extraction with phenol/chloroform. DNA was then precipitated with ethanol and resolved on a sequencing gel cotaining 10% polyacrylamide.

Kinetics of the Methyl Transfer Reaction. The kinetics of methyl group transfer were determined by monitoring the amount of [8-^{3}H]guanine generated in a synthetic substrate, poly(dC·dG·[8-^{3}H]m^{6}dG), as described earlier (Foote & Mitra, 1984). The initial concentration of the active methyltransferase was determined from a separate experiment in which an aliquot of the protein preparation was allowed to react with a 5-fold excess of substrate over an extended period of time (60 min) until the reaction was complete. Because the methyltransferase reaction is irreversible, the amount of the methylated protein corresponded to the initial amount of the active methyltransferase. In all measurements, aliquots of reaction mixtures were taken out at various times and the reaction was stopped by adding Sarkosyl NL97 to 0.5% before vortexing the solution with a mixture of phenol and chloroform (1:1) (Boulden et al., 1987). The reaction rate constant was calculated from the second-order rate equation (Frost & Pearson, 1961).

Other Methods. Proteins were quantitated by the bicinchoninic acid procedure with bovine serum albumin as the standard (Smith et al., 1985). Radioactivity in liquid samples was quantitated in a Beckman liquid scintillation counter. Oligodeoxynucleotides were synthesized in an Applied Biosystems model 394 DNA/RNA synthesizer.

RESULTS

Specific, Transient Complex between WT MGMT and Substrate Oligonucleotide. We were able to demonstrate, for the first time, the specific and reproducible binding of purified wild type hMGMT to an m^{6}G-containing oligonucleotide (Figure 1). This binding presumably represents the protein-substrate complex formed as an intermediate prior to the alkyl transfer reaction. Figure 1A shows the binding of full-length of hMGMT to a 27-mer oligonucleotide probe containing m^{6}G after incubation for 10 min at 20 °C before gel loading. This binding was specific, because the complex was effectively competed with by a 10-fold excess (25 pmol) of unlabeled (m^{6}G) oligonuleotide, but not by the same amount of unlabeled control oligonucleotide. To further confirm the specificity, EMSA was performed in the presence of an anti-hMGMT antibody. The antibody interacted with hMGMT and formed a more slowly migrating "supershifted" complex with m^{6}G oligonucleotide. In contrast, either preimmune serum or antibody in the absence of hMGMT did not cause any supershift of the hMGMT-bound complex (Figure 1A). Further evidence for specificity was the lack of binding of inactive truncated mutant MGMTs, NΔ7CΔ32 and NΔ9CΔ31 (Figure 1A). The properties of these mutants are described later.

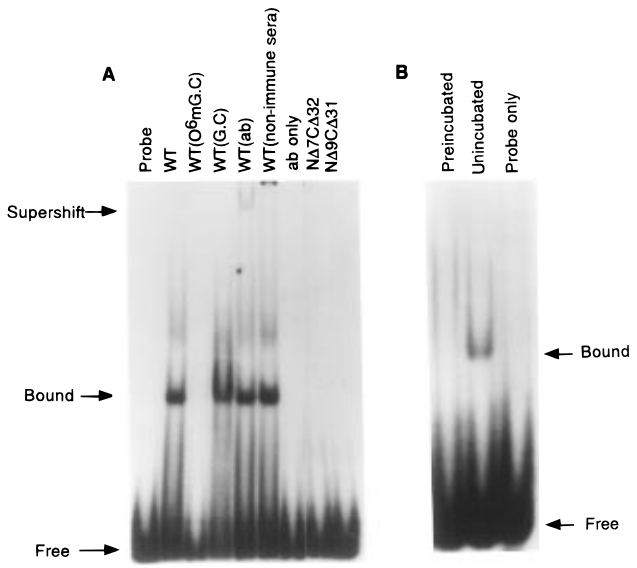

FIGURE 1: Gel mobility shift analysis of wild type and mutant MGMTs' binding to a 27-mer m^{6}G-containing duplex oligonucleotide. Unlabeled duplex oligonucleotides containing either m^{6}G or G, paired with C opposite m^{6}G or G, were used as competitors (in parentheses) (ab, anti-hMGMT antibody). The positions of free (free) and protein-bound DNA complex (bound) are indicated. (A) Specific substrate binding to wild type (WT) and not to inactive truncated mutant (NΔ7CΔ32 and NΔ9CΔ31) proteins. (B) Binding of m^{6}G-containing oligonucleotide to WT protein. Labeled m^{6}G-containing oligonucleotide (200−250 fmol) was mixed with 5.5 pmol of unlabeled oligonucleotide and incubated with 200 pmol of WT protein in the presence of 3 nmol of poly(dI·dC) in a 50 μL reaction volume at 37 °C for 30 min to remove the methyl group from m^{6}G. Eight to ten femtomoles of repaired oligonucleotide was used for gel shift analysis as described in Materials and Methods.

In order to establish that the wild type MGMT complex with the substrate was a transient reaction intermediate, labeled m^{6}G oligonucleotide was incubated with wild type hMGMT at 37 °C for 30 min prior to EMSA, and no complex was formed (Figure 1B). Furthermore, a 16 bp oligonucleotide fragment was released after digestion of the incubation mixture with *Bam*HI (data not shown) as expected from restoration of sensitivity to *Bam*HI following demeth-

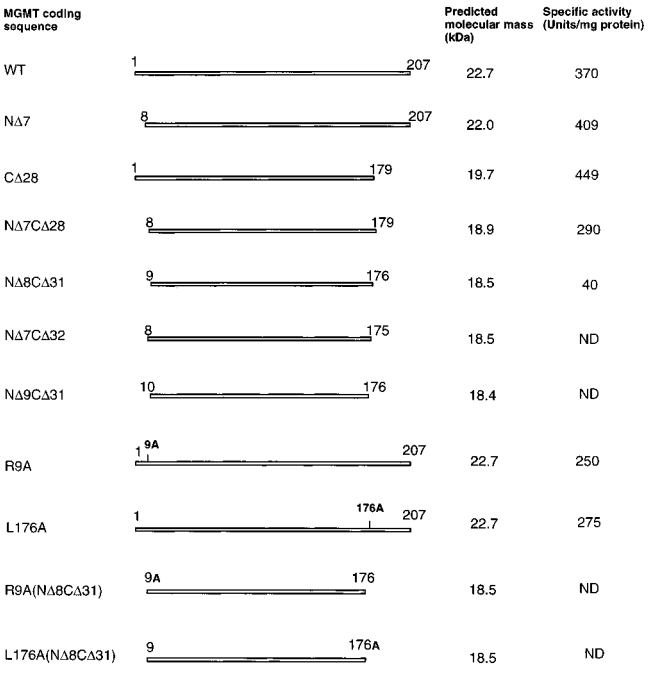

FIGURE 2: Schematic diagram of active, inactive deletion, and point mutants of hMGMT. The numbers of the first and last amino acid residues of hMGMT are shown at the ends of the open boxes; numbers in the middle indicate the site of mutagenesis. Predicted molecular masses and specific activities of the mutant proteins are also given. Specific activities of crude cell lysates were determined in three or four experiments, and the mean values are listed. ND means not detectable.

ylation of the m^6G, located at the 17th position of the oligonucleotide.

Analysis of Deletion Mutants of MGMT. A series of deletion constructs (Figure 2) were made in the human MGMT cDNA, and the proteins were expressed in *E. coli* BL21 (DE3) pLysS. The cell-free extracts were used to test hMGMT activity; the specific activities of different plasmid-encoded mutant MGMT after correcting for the background activity are shown in Figure 2. It is evident that 8 and 31 amino acid residues could be deleted from the N and C terminus, respectively, without concomitant loss of activity of the protein. It should be noted that the low specific activity (in whole cell lysate) of the NΔ8CΔ31 mutant protein in *E. coli* extract was due to its low level of expression. Removal of an additional residue from either end, i.e. Arg-9 at the N terminus and Leu-176 at the C terminus, reduced the hMGMT activity below the detectable level.

In order to test the possibility that Arg-9 and Leu-176 participate directly in the methyl transfer reaction, site-specific mutants R9A and L176A were generated in full-length and minimal-length constructs and expressed in *E. coli*. The bacterial cell-free extracts containing R9A or L176A in the full-length protein had specific activity comparable to that with the full-length wild type protein. In contrast, cell-free extracts of the same mutants in minimal-length (NΔ8CΔ31) hMGMT had no detectable activity (Figure 2). This observation suggests that neither Arg-9 nor Leu-176 is essential for MGMT activity.

Expression of Wild Type and Mutant MGMTs. Because the lack of hMGMT activity in cell-free extracts of *E. coli* harboring mutant plasmids may be due to poor expression of the mutant proteins, it was important to check the level of expression of these proteins. Western blot analysis (Figure 3) of the total soluble protein showed that NΔ8CΔ31 and NΔ9CΔ31 mutant proteins, as well as R9A and L176A mutants in truncated form, were expressed poorly in *E. coli*, whereas wild type and NΔ7, CΔ28, R9A, and L176A mutants had high levels of expression as full-length proteins.

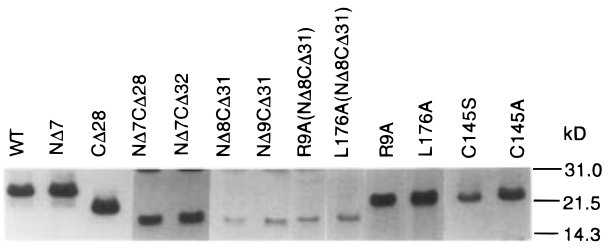

FIGURE 3: Expression of deletion mutants in *E. coli*, analyzed by Western blotting. Blots of *E. coli* extracts containing hMGMT mutants were probed with an anti-hMGMT antibody. The positions of rainbow protein markers (14.3−31 kDa, Amersham) are indicated.

Table 1: Second-Order Rate Constants and Association Constants of WT, Inactive Full-Length, and Truncated MGMTs

protein	rate constant (M^{-1} min^{-1})	association constant (M^{-1})	
		specific	nonspecific
wild type	5.2×10^{-7}	−	3.8×10^5
CΔ28	6.1×10^7	−	−
NΔ7CΔ28	1.8×10^7	−	−
C145A	−	1.3×10^6	4.2×10^5
C145S	−	0.8×10^6	3.1×10^5

Kinetics of the Methyl Transfer Reaction. The second-order rate constant of the NΔ7CΔ28 mutant was approximately 3-fold lower than that of the CΔ28 deletion mutant, which reacted at a rate comparable to that of the wild type protein (Table 1). These data suggest that a small structural alteration was induced in NΔ7CΔ28 but not in the CΔ28 truncated mutant. The CΔ28 mutant may be appropriate for NMR studies because of its size, activity, and high level of expression in soluble form in *E. coli* that will facilitate its large scale purification.

Stable Complexes of Substrate and Active Site Mutants. Site-specific mutation at the conserved active site (C145S and C145A) in full-length hMGMT completely eliminated the activity, as expected. However, both inactive mutants retained specific binding activity to substrate oligonucleotide, and the complexes were stable even at 37 °C, unlike in the case of wild type hMGMT (Figure 4). In addition, wild type and C145S and C145A mutants showed comparable non-specific binding to a control duplex oligonucleotide having the same sequence as the 27-mer m^6G oligonucleotide duplex except for the replacement of m^6G with guanine (Figure 4B).

The binding constants of C145A and C145S mutant proteins for the substrate m^6G-containing oligonucleotide were comparable (Table 1). Although these were apparently about 3-fold higher than that for the binding of wild type and the mutant MGMT to control oligonucleotide, these values cannot be compared because the specific binding assay, unlike for nonspecific DNA binding, was carried out in the presence of a very large excess of competitor poly-(dI·dC). The wild type hMGMT and site-specific mutants showed comparable nonspecific binding affinity to control oligonucleotide, suggesting that the mutations did not cause a major structural change in the protein (Table 1). However,

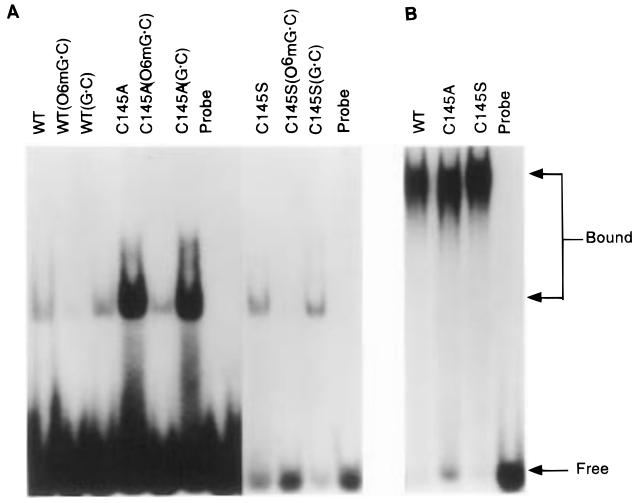

FIGURE 4: (A) Specific binding of purified WT, C145A, and C145S proteins at 37 °C. The oligonucleotide competitors were as described in Figure 1. Other details are given in Materials and Methods. (B) Nonspecific binding of WT, C145A, and C145S proteins to a 27-mer control oligonucleotide. Each reaction mixture (15 μL) contained 8−10 fmol of ^{32}P-labeled oligonucleotide, 10 pmol of unlabeled oligonucleotide, and 220 pmol of pure protein in the absence of poly(dI·dC). After 10 min at 37 °C, the nucleoprotein complexes were resolved in a 6% nondenaturing polyacrylamide gel.

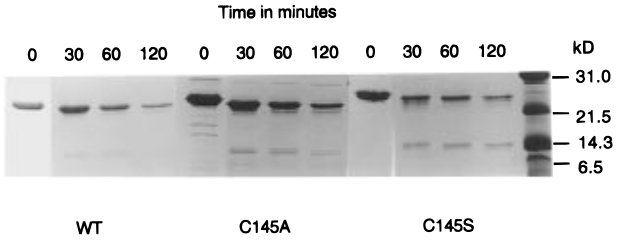

FIGURE 5: SDS−PAGE assay for trypsin sensitivity of hMGMT. Five micrograms of WT, C145A, and C145S mutant proteins were digested with 0.05 μg of trypsin and aliquots taken out at indicated times for SDS−PAGE. Other details are described in Materials and Methods.

the specific binding affinity constant could not be determined for the wild type protein, due to its rapid reaction at 37 °C with m^{6}G in the oligonucleotide.

Similar Protease Sensitivity of Mutant and Wild Type MGMTs. Limited proteolysis was used as a probe to detect any major structural alteration in C145S and C145A mutants as a result of mutation. After treatment of wild type hMGMT and C145S and C145A mutants with trypsin at room temperature for up to 2 h, three major peptide fragments of discrete and identical sizes were seen with all three proteins (Figure 5). Although the C145A mutant preparation contained some minor contamination of smaller peptides, this result strongly suggests that conservative substitution of Cys-145 with Ser or Ala did not cause any major change in the tertiary structure of hMGMT.

Footprinting Analysis of C145A MGMT-Bound Substrate. The DNase I footprint for C145A mutant proteins bound to m^{6}G oligonucleotide is shown in Figure 6. When the m^{6}G-containing strand was labeled, a footprint of 17−18 nucleotides was observed spanning the m^{6}G site. Seven nucleotides on the 5′ side of m^{6}G and 10 nucleotides on the 3′ side of m^{6}G were protected from DNase digestion (Figure 6B). But when the complementary strand was labeled, a footprint of 13 nucleotides was observed (Figure 6C).

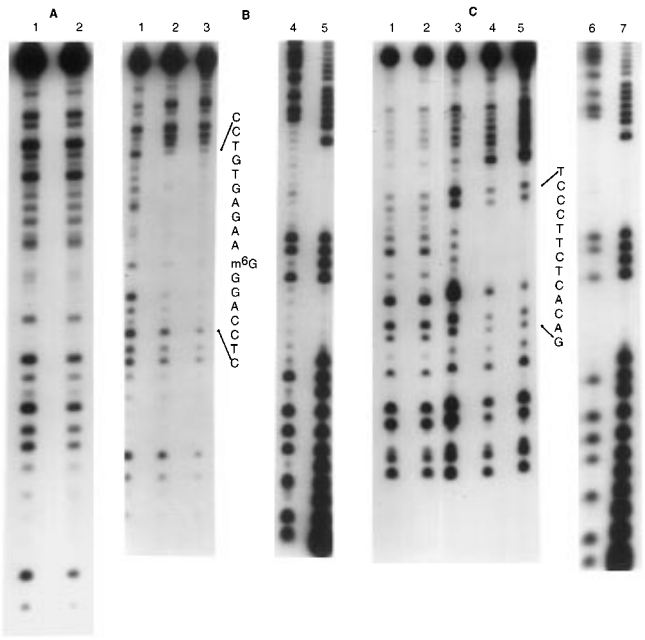

FIGURE 6: DNase I footprinting of the MGMT−substrate DNA complex. Oligonucleotides (45-mer) containing m^{6}G in the substrate or G in the control DNA at the 22nd position were annealed with the same complementary strand after labeling one of the strands with ^{32}P at the 5′ end, and the DNA was incubated with 25−50 pmol of C145A mutant before DNase I digestion and subsequent denaturing PAGE as described in Materials and Methods. (A) Lane 1, ^{32}P-labeled control DNA alone; and lane 2, after incubation with 50 pmol of protein. (B) ^{32}P-labeled m^{6}G-containing strand. Lanes 1−3, with no protein and 25 and 50 pmol of protein, respectively. Lanes 4 and 5 represent G and G + A Maxam−Gilbert sequencing reactions, respectively. (C) ^{32}P-labeled complementary strand in control (lanes 1 and 2) or m^{6}G-containing duplex (lanes 3−5). Lane 1, no protein; lane 2, 50 pmol of protein; and lanes 3−5, 0, 25, and 50 pmol of protein, respectively. Lanes 6 and 7 represent G and G + A Maxam−Gilbert sequencing reactions, respectively.

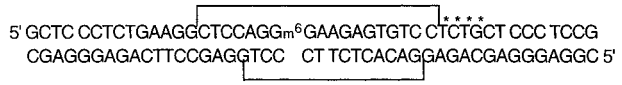

FIGURE 7: Site-specific binding of C145A MGMT to m^{6}G-containing sequence as obtained from DNase I footprinting shown in Figure 6. Protected residues are bracketed. Residues hypersensitive to DNase I are marked with asterisks.

Protection from DNase I was definitive on both strands, but the terminal nucleotides on both 5′ and 3′ ends were considerably less protected. No DNase I footprint was observed for the unmodified oligonucleotide, a further indication that the binding of C145A to m^{6}G oligonucleotide is a specific interaction. Figure 7 shows the oligonucleotide region protected by C145A MGMT.

DISCUSSION

The ubiquitous nature of MGMT, a highly unusual protein because of its stoichiometric alkyl transfer reaction, underscores its importance in removing endogenous and induced *O*^{6}-alkylguanine residues from DNA. The sequence of MGMT is not globally conserved among bacteria, yeast, and mammals, except in distinct regions of the protein, in particular that surrounding the alkyl acceptor cysteine residue (Moore et al., 1994). Thus, although the structure of the 19 kDa MGMT domain of the *E. coli* Ada protein was recently solved by X-ray crystallography, it is important to elucidate the structure and reaction mechanism of mammalian MGMTs

because of their distinct biochemical properties relative to the bacterial protein (Dolan et al., 1991).

Using EMSA, we demonstrated for the first time that wild type hMGMT forms a specific complex with m^6G-containing oligonucleotide at 20 °C, because the binding was not competed for by a G-containing oligonucleotide having the same sequence as the substrate. Furthermore, the complex appears to be a transient intermediate for the methyl transfer reaction because preincubation of the substrate with MGMT abolished the binding. Such a transient intermediate is obviously not suitable for structural studies of the MGMT−substrate complex.

We thus decided to identify the stable and specific dead-end complex of m^6G-containing DNA with inactive MGMT mutants that should have minimal structural perturbation as a result of mutation. We used two approaches to generate such mutants, namely, deletion mutagenesis and conservative substitution of the active site residue. The rationale for isolating deletion mutants of MGMT that retained specific substrate binding was that small size as well as stability and high solubility of a protein are critical factors for structural determination by NMR spectroscopy. We have recently been successful in isolating inactive truncated mutants of mouse *N*-methylpurine-DNA glycosylase, another key alkylation adduct repair protein, which binds to the substrate DNA (R. Roy et al., manuscript in preparation). Although earlier studies showed that hMGMT retains activity after deletion of 31 amino acid residues from the C terminal (Crone et al., 1994), no systematic studies were carried out to determine the minimum size needed for activity by deleting residues from both termini. We have now shown that 8 and 31 residues can be deleted from the N and C terminus of full-length MGMT, respectively, but deletion of one additional residue from either end abolished activity as well as substrate binding. It is also surprising that the level of expression of mutant MGMTs in *E. coli* was strongly affected by a small change at the N terminus. For example, expression of both NΔ7CΔ28 and NΔ7CΔ32 mutant proteins is comparable to that of the wild type hMGMT. However, removal of one or more residues from the N terminus caused a drastic drop in expression of the truncated proteins. Furthermore, CΔ28 had reaction kinetics 3-fold faster than those of NΔ7CΔ28. Thus, on the basis of expression level, activity, and solubility of the recombinant protein in *E. coli*, the 19 kDa CΔ28 mutant, while not the minimum size active protein, is the protein of choice for structural studies by NMR.

The observation that truncated mutants, lacking Arg-9 and Leu-176 were inactive raised the possibility that these residues may be involved in the alkyl transfer reaction. However, that possibility was precluded by the observation that mutations of these residues to Ala in the full-length protein did not affect activity. Because the lack of activity of truncated mutants was not due to poor expression in *E. coli*, we concluded that Arg-9 and Leu-176 residues are involved in maintaining the structural integrity of the protein. While their contribution is not essential in the full-length protein, it becomes critical in the absence of distal residues. Crone et al. (1996) have recently shown that the WT protein and C145A mutant are quite stable when expressed in *E. coli* (half-life of >12 h), whereas mutants lacking 10 or 19 residues from the amino end are unstable, with half-lives of 90 and 48 min, respectively. This result is consistent with our conclusion about the structural alteration in the mutants lacking 9 or more residues from the amino end of the protein.

Our second approach was to examine whether active site mutants retain specific substrate recognition, and thus, their interaction with the substrate can be studied by NMR spectroscopy and X-ray crystallography. We showed earlier that the C145A mutant binds nonspecifically to DNA (Kanugula et al., 1995). We have now established that the inactive C145A and C145S mutants also retain specific substrate recognition. Thus, unlike the WT protein, their specific complexes with the substrate are stable at 37 °C. This result suggests that conservative substitution of the active site residue did not cause a significant alteration of the protein's tertiary structure. This conclusion was further supported by the similar protease sensitivity of the WT and mutant proteins. Although both C145A and C145S mutants are expressed in soluble form in *E. coli*, expression of the Ser-145 mutant was consistently lower than that of the Ala-145 or of wild type MGMT. Therefore, the C145A mutant may be appropriate for elucidating the structure of both free and substrate-bound MGMT. Whether the C145A, CΔ28 double mutant will provide a further advantage for NMR spectroscopic analysis remains to be tested. In this context, we should point out that other residues such as Arg-128 and Tyr-114 were also shown to be critical for MGMT activity (Kanugula et al., 1995); however, we have not tested whether the inactive mutants at these sites also form specific complexes with m^6G-containing DNA.

The final evidence for specific DNA interaction of the C145A mutant was provided by the footprinting studies (Figure 6) which showed that MGMT prevents DNase I digestion of a defined number of bases, spanning m^6G in a duplex substrate DNA. Comparison of lanes 1 and 3 in panel B of Figure 6 suggests higher DNase I sensitivity of residues TCTG 3′ to the protected region in the MGMT−DNA complex than in the control. This preliminary result is consistent with the expectation that protein binding induced a twist in the m^6G strand beyond the bound region, resulting in increased exposure to DNase. Elucidation of the tertiary structure of the MGMT−DNA compex will identify interactions between specific amino acid residues with the bases. We stress that this protection from DNase was not due to sequence-specific binding of MGMT, whose recognition of the substrate lesion should not be sequence-dependent. Although we have not analyzed footprinting of other sequences, the fact that we could detect no protection of the control oligonucleotide having the identical sequence except for substitution of m^6G with G strongly argues against MGMT recognition of a specific sequence. Furthermore, a different oligonucleotide sequence was used to show interaction with MGMT in EMSA (Figures 1 and 4).

The affinity of MGMT for substrate DNA was calculated earlier to be only about 3-fold higher than that for nonspecific DNA (Chan et al., 1993). Although our results appear to support this observation (Table 1), a comparison of the affinity constants for specific and nonspecific DNA binding in our studies is inappropriate. This is because binding to the substrate oligonucleotide was allowed in the presence of a large excess of competitor DNA [poly(dI·dC)], while nonspecific binding to the control oligonucleotide was carried out in the absence of the competitor. Thus, the initial concentration of free MGMT in substrate-specific binding was far less than the nominal value because a large fraction

of the protein was tied up by poly(dI·dC). Furthermore, our recent results indicate that nonspecific binding of MGMT to DNA may be cooperative with oligomeric protein in contrast to equimolar specific binding (Fried et al., 1996). Multimeric binding of the protein to control DNA was also suggested in our studies by the mobility of the nonspecific DNA complex in EMSA being lower than that of the protein−substrate complex (Figure 4B). Finally, the specificity of the C145A mutant in substrate binding, as indicated by DNA footprinting studies, makes it an excellent surrogate for studying the structure and reaction mechanism of human MGMT.

ACKNOWLEDGMENT

We thank Drs. David Konkel and Gourisankar Ghosh for a critical reading of the manuscript and Ms. Wanda Smith for expert secretarial assistance.

REFERENCES

Bhattacharyya, D., Foote, R. S., Boulden, A. M., & Mitra, S. (1990) *Eur. J. Biochem. 193*, 337−343.
Boulden, A. M., Foote, R. S., Flemming, G. S., & Mitra, S. (1987) *J. Biosci. 11*, 215−224.
Brent, T. P., Houghton, J. A., & Houghton, P. J. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2985−2989.
Chan, C.-L., Wu, Z., Ciardelli, T., Eastman, A., & Bresnick, E. (1993) *Arch. Biochem. Biophys. 300*, 193−200.
Crone, T. M., Goodtzova, K., Edara, S., & Pegg, A. E. (1994) *Cancer Res. 54*, 6221−6227.
Crone, T. M., Goodtzova, K., & Pegg, A. E. (1996) *Mutat. Res. 363*, 15−26.
Demple, B., Jacobson, A., Olsson, M., Robins, P., & Lindahl, T. (1982) *J. Biol. Chem. 257*, 13776−13780.
Deng, W. P., & Nickoloff, J. A. (1992) *Anal. Biochem. 200*, 81.
Dolan, M. E., Pegg, A. E., Dumenco, L. L., Moschel, R. C., & Gerson, S. L. (1991) *Carcinogenesis 12*, 2305−2309.
Dosanjh, M. K., Roy, R., Mitra, S., & Singer, B. (1994) *Biochemistry 33*, 1624−1628.
Foote, R. S., & Mitra, S. (1984) *Carcinogenesis 5*, 277−281.
Foote, R. S., Mitra, S., & Pal, B. C. (1980) *Biochem. Biophys. Res. Commun. 97*, 654−659.
Fried, M. G., Kanugula, S., Bromberg, J. L., & Pegg, A. E. (1996) *Biochemistry 35*, 15295−15301.
Frost, A. A., & Pearson, R. S. (1961) *Kinetics and Mechanisms: A Study of homogeneous chemical reaction*, 2nd ed., p 15, John Wiley & Sons, New York.
Hayakawa, H., Koike, G., & Sekiguchi, M. (1990) *J. Mol. Biol. 213*, 739−747.
Kanugula, S., Goodtzova, K., Edara, S., & Pegg, A. E. (1995) *Biochemistry 34*, 7113−7119.
Laemli, U. K. (1970) *Nature 277*, 680−685.
Leblanc, B., & Moss, T. (1994) *Methods in Molecular Biology 30* (Kneale, G. G., Ed.) pp 1−10, Humana Press, Inc., Totowa, NJ.
Lindahl, T., Sedgwick, B., Sekiguchi, M., & Nakabeppu, Y. (1988) *Annu. Rev. Biochem. 57*, 133−157.
Loechler, E. L., Green, C. L., & Essigman, J. M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 6271−6275.
Loveless, A. (1969) *Nature 223*, 206−207.
Mitra, S., & Kaina, B. (1993) *Prog. Nucleic Acids Res. Mol. Biol. 44*, 109−141.
Moore, M. H., Gulbus, J. H., Dobson, E. J., Demple, B., & Moody, P. C. E. (1994) *EMBO J. 13*, 1495−1501.
Morgan, S. E., Kelley, M. R., & Pieper, R. O. (1993) *J. Biol. Chem. 268*, 19802−19809.
Olsson, M., & Lindahl, T. (1980) *J. Biol. Chem. 255*, 10569−10571.
Pegg, A. E. (1990) *Cancer Res. 50*, 6119−6129.
Potter, P. M., Wilkinson, M. C., Fitton, J., Carr, F. J., Brennand, J., Copper, D. P., & Margison, G. P. (1987) *Nucleic Acids Res. 15*, 9177−9193.
Rebeck, G. W., Coons, S., Carrol, P., & Samson, L. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 3039−3043.
Roy, R., Shiota, S., Kennel, S. J., Raha, R., von Wronski, M., Brent, T. P., & Mitra, S. (1995) *Carcinogenesis 16*, 405−411.
Shiota, S., von Wronski, M. A., Tano, K., Bigner, D. D., Brent, T. P., & Mitra, S. (1992) *Biochemistry 31*, 1897−1903.
Singer, B., & Grunberger, D. (1983) *Molecular Biology of Mutagens and Carcinogens*, Plenum Press, New York.
Smith, P. R., Krohn, P. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Porvenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., & Kleuk, D. C. (1985) *Anal. Biochem. 150*, 76−85.
Snow, E. T., Foote, R. S., & Mitra, S. (1984) *J. Biol. Chem. 259*, 8095−8100.
Stone, S. R., Hughes, M. J., & Jost, J. P. (1991) Qualitative and Quantitative Studies of Protein-DNA Interactions by Gel Mobility-Shift Assay, in *BioMethods 5*, pp 163−184, Birkhauser Verlag, Basel.
Tano, K., Shiota, S., Collier, J., Foote, R. S., & Mitra, S. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 686−690.
von Wronski, M. A., Shiota, S., Tano, K., Mitra, S., Bigner, D. D., & Brent, T. P. (1991) *J. Biol. Chem. 266*, 1064−1070.
Waldstein, E. A., Cao, E. H., & Setlow, R. B. (1982) *Nucleic Acids Res. 10*, 4595−4604.
Wüthrich, K. (1996) in *Encyclopedia of Nuclear Magnetic Resonance* (Grant, D. M., & Davis, R. K., Eds.) Vol. 2, p 939, John Wiley, New York.

BI963085I