Biochemistry

 Volume 31, Number 7 February 25, 1992

Articles

Characterization of cDNA Encoding Mouse DNA Repair Protein O^6-Methylguanine–DNA Methyltransferase and High-Level Expression of the Wild-Type and Mutant Proteins in *Escherichia coli*[†,‡]

Susumu Shiota,[§] Mathew A. von Wronski,[∥] Keizo Tano,[§] Darell D. Bigner,[⊥] Thomas P. Brent,[∥] and Sankar Mitra*[,#]

University of Tennessee—Oak Ridge Graduate School of Biomedical Sciences and Biology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, Department of Pharmacology, St. Jude Children's Research Hospital, Memphis, Tennessee 38101, and Department of Pathology, Duke University Medical Center, Durham, North Carolina 27710

Received August 7, 1991; Revised Manuscript Received November 22, 1991

ABSTRACT: A mouse cDNA clone encoding O^6-methylguanine–DNA methyltransferase (MGMT), responsible for repair of mutagenic O^6-alkylguanine in DNA, was cloned from a λgt11 library. On the basis of an open reading frame in cDNA, the mouse protein contains 211 amino acids with a molecular mass of 22 kDa. The size and the predicted N-terminal sequence of the mouse protein were confirmed experimentally. The deduced amino acid sequence of the mouse MGMT is 70% homologous to that of the human MGMT. Cysteine-149 was shown to be the only alkyl acceptor residue in the mouse protein, in confirmation of the prediction based on conserved sequences of different MGMTs. Mouse MGMT protein is recognized by some monoclonal antibodies specific for human MGMT. Site-directed mutagenesis was utilized to reclone the mouse cDNA in a T7 promoter-based vector for overexpression of the native repair protein in *Escherichia coli*. The mouse protein has a tetrapeptide sequence, Pro-Glu-Gly-Val at positions 56–59, absent in the human protein. Neither deletion of this tetrapeptide nor substitution of valine-169 with alanine affected the activity of the mutant proteins.

Simple alkylating mutagens induce a number of base adducts in DNA among which O^6-alkylguanine is the critical mutagenic and carcinogenic lesion (Loveless, 1969; Pegg & Singer, 1984). This adduct is removed by a highly unusual repair protein, O^6-methylguanine–DNA methyltransferase (MGMT) (EC 2.1.1.63), first discovered in *Escherichia coli* (Foote et al., 1980; Olsson & Lindahl, 1980). MGMT removes O^6-alkylguanine by accepting the alkyl group at a unique cysteine residue in a stoichiometric, second-order reaction (Bhattacharyya et al., 1990; Bogden et al., 1981; Margison et al., 1985; Olsson & Lindahl, 1980; von Wronski et al., 1991) that results in its inactivation and restoration of the original guanine residue. MGMT activity is present in all organisms tested so far (Lindahl et al., 1988; Pegg, 1990).

The bacterial MGMTs have been extensively characterized, and the genes have been cloned from *E. coli* and *Bacillus subtilis*. *E. coli* has two distinct genes that encode MGMT proteins. The *ada* gene encodes the inducible 39-kDa Ada protein that regulates its own and other alkylation repair genes (Demple et al., 1985; Nakabeppu & Sekiguchi, 1986). The second MGMT gene of *E. coli*, *ogt*, encodes the noninducible 19-kDa Ogt protein expressed at a low basal level (Potter et al., 1987; Rebeck et al., 1988). In *B. subtilis*, *dat*-1, which encodes a constitutive MGMT, has been cloned and sequenced (Kodama et al., 1989; Morohoshi et al., 1989). The predicted

[†] This work was supported by National Institutes of Health Grant CA 31721 and by the Office of Health and Environmental Research, U.S. Department of Energy, under Contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc. (to Oak Ridge National Laboratory), by National Institutes of Health Grants CA 11898 and NS 20023 and Bristol-Myers Squibb Grant 100-R18 (to Duke University Medical Center), and by National Institutes of Health Grants CA 14799, CA 36888, CA 09346, and CA 21765 and American Lebanese Syrian Associated Charities (to St. Jude Children's Research Hospital).

[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J05360.

* Address correspondence to this author.

[§] University of Tennessee—Oak Ridge Graduate School of Biomedical Sciences.

[∥] St. Jude Children's Research Hospital.

[⊥] Duke University Medical Center.

[#] Oak Ridge National Laboratory.

0006-2960/92/0431-1897$03.00/0

amino acid sequence of this protein shows striking similarity to those of *E. coli* MGMTs.

The level of MGMT in mammalian cells and tissues appears to be highly regulated. Since the original observation of a large difference in repair activity for O^6-alkylguanine in brain and liver of rodents some 2 decades ago, MGMT activity was found to be tissue- and cell-type-specific and varies with cell cycle and age in rodents (D'Ambrosio et al., 1987; Washington et al., 1989). Furthermore, many cell lines of both human and rodent origin, called Mer^-/Mex^-, have no detectable MGMT activity (Day et al., 1980; Sklar & Strauss, 1981). Several established hamster cell lines (e.g., V79 and CHO) and 20–30% of human tumor lines are Mer^- (Foote & Mitra, 1984; Pegg, 1990). At the same time, Mer^+ tumor cells vary widely in their resistance to alkylating drugs such as procarbazine and 2-chloroethyl-*N*-nitrosourea derivatives, and the degree of resistance is directly related to their MGMT levels (Brent et al., 1985; Schold et al., 1989). The molecular mechanism of the regulation of MGMT expression is not clear. In order to elucidate the basis of such regulation, we recently cloned a cDNA of human MGMT (Tano et al., 1990). Similar cDNAs were isolated later by other laboratories by using different techniques (Hayakawa et al., 1990; Rydberg et al., 1990). The amino acid sequence of human MGMT protein deduced from the nucleotide sequence has shown significant homology with the C-terminal half of *E. coli* Ada protein. In particular, the cysteine residue in the sequence—PCHRV—that is completely conserved in the *E. coli* and the human proteins has been shown to be the alkyl acceptor site (von Wronski et al., 1991).

Even though the mammalian MGMTs appear to have similar reaction conditions and substrate specificity, their reaction parameters have not been studied to a significant extent because of the lack of availability of purified proteins. A comparative study of the physicochemical and biochemical properties of the wild-type and mutant MGMTs may elucidate the precise reaction mechanism of this unusual class of proteins. Furthermore, the information on the structure and organization of the mouse *Mgmt* gene would provide an opportunity for introducing a null mutation in the gene in the animal by targeted mutagenesis for investigating the role of this protein in vivo. As a first step in this study, this report describes isolation and characterization of a mouse MGMT cDNA and construction of a high-level expression vector for large-scale production of wild-type and mutant mouse MGMTs.

Materials and Methods

Bacterial Strain. Escherichia coli JM109, which carries *lacI*^{q}, was used for expression of mouse MGMT, and *E. coli* BL21(DE3) (Studier et al., 1990) was used as the host for high-expression plasmid.

Assay of Methyltransferase. Synthetic poly(dC, dG, [8-^{3}H]m^{6}dG) was used as the substrate for the MGMT assay as described (Foote & Mitra, 1984). The growth of *E. coli* cells and preparation of crude extracts used in these assays have been described (Tano et al., 1988).

Nucleotide Sequence Analysis. The cDNA insert of pSS601 was sequenced in both directions by the "dideoxy" chain-termination method (Sanger et al., 1977) with Sequenase version 2.0 (United States Biochemical) and [^{35}S]dATP on double-stranded templates. pSS601 was used as the template first with oligodeoxynucleotide primers of sequences corresponding to the regions of pUC18 plasmid flanking the cDNA insert. Sequencing in the later rounds was carried out with primers whose sequences were derived from the appropriate regions of the cDNA sequenced in the previous round.

Nucleic Acid Hybridization. Ten micrograms of DNA was digested with *Eco*RI, separated by electrophoresis on an 0.8% agarose gel, and finally transferred to nylon membrane (Hybond-N, Amersham) for hybridization (Maniatis et al., 1982) with mouse *Mgmt* cDNA. The cDNA probe was labeled with [^{32}P]dCTP by the random hexanucleotide priming method (Feinberg & Vogelstein, 1983). Hybridization was carried out at 65° for 20 h in 0.5 M sodium phosphate butter (pH 7.2) containing 2 mM EDTA, 7% SDS, and 1% nonfat dry milk. The filter was washed with 0.3 M NaCl/0.03 M sodium citrate, pH 7.4, and 0.1% SDS at 65 °C for 1 h, dried, and exposed to Kodak X-ray film.

For RNA blot analysis, total RNAs were prepared from cultured cells by a modified guanidinium thiocyanate procedure (Chomczynski & Sacchi, 1987). Twenty micrograms of total RNA was separated by electrophoresis on a 1.2% agarose gel containing 2.2 M formaldehyde followed by capillary transfer to a nylon membrane (Hybond-N, Amersham). Hybridization with the ^{32}-P-labeled cDNA insert of pSS601 was carried out as described above. The filters were washed with 0.15 M NaCl/0.015 M sodium citrate, pH 7.0, and 0.1% SDS for 1 h at 65 °C and exposed to X-ray film.

Primer Synthesis and Polymerase Chain Reaction (PCR) Protocol. Synthetic oligonucleotide primers which were complementary to a segment of the translation initiation region were prepared in an oligonucleotide (PCR Mate, Applied Biosystems, Inc.) synthesizer. The primers were designed so as to be "back to back" on the duplex. One primer had the same sequence as the sense strand except for three bases just before the start codon for creation of a *Nde*I recognition site, and another primer was exactly the same with the opposite strand started from the neighbor base of the 5′ end of the first primer (5′CGGGGGCCATATGGCTGA3′ and 5′AATTCGTAATCATGGTCAT3′).

The PCR procedures for oligonucleotide primer-mediated site-directed mutagenesis were carried out as described by Hemsley et al. (1989).

Purification of Mouse MGMT. The MGMT encoded by the mouse cDNA cloned in a T7 promoter-based expression vector was purified from plasmid-containing *E. coli* BL21-(DE3) (Studier et al., 1990) in essentially the same way as described for the recombinant human MGMT expressed in *E. coli* (von Wronski et al., 1991). [CH_3-^{3}H]-methylated MGMT was prepared by incubating the pure protein with [CH_3-^{3}H]-*N*-methyl-*N*-nitrosourea-treated DNA as described earlier (Brent, 1985).

SDS–Polyacrylamide Gel Electrophoresis. Proteins were separated by SDS–PAGE in 0.75-mm-thick slab gels according to the method of Laemmli (1970) in a Bio-Rad minigel apparatus at 200 V for 45 min. Bio-Rad low molecular mass standards (rabbit muscle phosphorylase *b*, 97 kDa; bovine serum albumin, 66 kDa; hen egg white ovalbumin, 42.7 kDa; bovine carbonic anhydrase, 31 kDa; soybean trypsin inhibitor, 21.5 kDa; hen egg white lysozyme, 14.4 kDa) were used to calibrate the gels.

Electroblotting and Autoradiography. Proteins were transferred from gels onto a PVDF membrane (Immobilon-P, Millipore, Bedford, MA) by the method of Matsudaira (1987) using a Bio-Rad Mini Trans-Blot cell, 1 h at 70 mA. Electroblots were autoradiographed directly on Kodak X-Omat AR film (Eastman Kodak Co., Rochester, NY) without prior enhancement for 72 h at −70 °C.

Isolation and Characterization of the S-Methylcysteine-Containing Peptide. Fifty micrograms of [CH_3-^{3}H]-methylated MGMT was acetone-precipitated (von Wronski

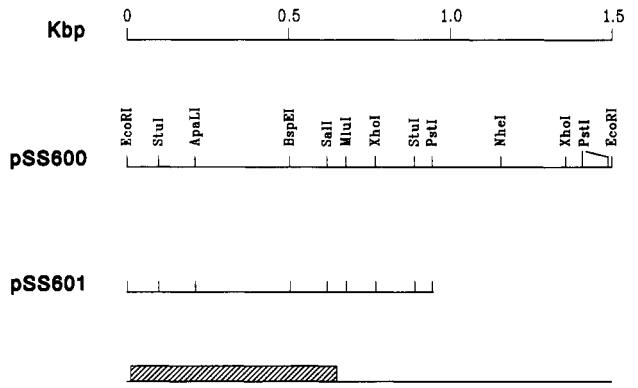

FIGURE 1: Physical map of the cDNA insert. The predicted open reading frame from the nucleotide sequence is indicated by the hatched box.

et al., 1991) and redissolved in 40 mM Tris-HCl (pH 8.95)/0.05% SDS. After being heated at 80 °C for 5 min, the solution was incubated with 0.5 μg of lysyl endopeptidase (Wako Chemicals, Richmond, VA) for 2.5 h at room temperature. Following addition of one-third volume of 4× electrophoresis buffer and heating at 100 °C for 3 min, the samples were loaded on a 15% SDS–polyacrylamide gel with Tricine instead of glycine in the running buffer (Schagger & von Jagow, 1987) for subsequent electroblotting and autoradiography. The radioactive peptide band was excised from the blotted membrane, and the amino acid sequence of the peptide was determined in an automated gas-phase sequenator. At the same time, the radioactivity of chromatographic fractions of amino acid derivatives was determined. These procedures have been reported earlier (von Wronski et al., 1991).

Immunoprobing of the Western Blot. Protein blotted onto Immobilon-P membrane was probed with monoclonal antibodies 4.A1, 19.2, and 21.A8 that are specific for human O^6-methylguanine–DNA methyltransferase (von Wronski et al., 1989). Antibody binding was visualized by a silver-enhanced, gold staining procedure (Brent et al., 1990).

RESULTS

Cloning and Nucleotide Sequence of Mouse MGMT cDNA. We have previously reported that the coding sequences among the human and rodent MGMT genes were conserved (Tano et al., 1990). We used the 835 bp long human MGMT cDNA containing *Eco*RI linkers at both ends as a probe for screening a λgt11 mouse cDNA library (Clontech). After we screened about 3.3×10^5 plaques of recombinant phage, one cDNA clone which hybridized to human MGMT cDNA was isolated. This recombinant phage contained a 1.5-kb *Eco*RI fragment, which was recloned on pUC9 (pSS600). The cDNA fragment is much larger than the 1-kb mouse MGMT mRNA. The physical map of the cDNA insert in pSS600 was established, and the entire coding sequence of mouse MGMT was found to be contained in a 0.95-kb *Eco*RI/*Pst*I fragment which was subcloned in pUC18 (pSS601) (Figure 1). The 0.95-kb inserted fragment of pSS601 was sequenced and was found to contain 946 bp (Figure 2).

Identification of the Open Reading Frame and Sequence Homology with the Human and Bacterial MGMTs. The cDNA sequence was examined for distribution of the potential stop codon in each of the six possible reading frames. There was one open reading frame which started at positions 16–18 (ATG) with a termination codon (TGA) at positions 848–850. This open reading frame would encode 211 amino acids with a molecular mass of 22 kDa. The MGMT was purified to

```
1                                                       GAATTCGGGGGCAGA
16  ATG GCT GAG ACC TGC AAA ATG AAA TAC TCA GTG TTG GAC AGC CCT TTG
     M   A   E   T   C   K   M   K   Y   S   V   L   D   S   P   L
     1                                   10
64  GGG AAG ATG GAG CTG TCT GGC TGT GAG CGA GGC CTG CAT GGG ATA CGG
     G   K   M   E   L   S   G   C   E   R   G   L   H   G   I   R
             20                                              30
112 TTG CTC AGT GGG AAG ACC CCA AAC ACT GAC CCC ACA GAG GCC CCA GCT
     L   L   S   G   K   T   P   N   T   D   P   T   E   A   P   A
                             40
160 ACT CCT GAG GTG CTC GGT GGG CCA GAG GGA GTT CCA GAG CCT CTG GTG
     T   P   E   V   L   G   G   P   E   G   V   P   E   P   L   V
         50                                          60
208 CAG TGC ACA GCC TGG CTG GAA GCC TAT TTC CGT GAA CCC GCA GCC ACA
     Q   C   T   A   W   L   E   A   Y   F   R   E   P   A   A   T
                 70                                          80
256 GAG GGG CTT CCC TTG CCT GCT CTC CAT CAC CCT GTG TTC CAG CAA GAT
     E   G   L   P   L   P   A   L   H   H   P   V   F   Q   Q   D
                                     90
304 TCA TTC ACC AGA CAG GTG TTA TGG AAG CTG CTG AAG GTT GTG AAA TTC
     S   F   T   R   Q   V   L   W   K   L   L   K   V   V   K   F
             100                                         110
352 GGA GAA ACG GTT TCT TAC CAG CAA TTA GCA GCC CTG GCA GGC AAC CCC
     G   E   T   V   S   Y   Q   Q   L   A   A   L   A   G   N   P
                         120
400 AAA GCG GCT CGT GCA GTA GGA GGA GCA ATG AGA AGC AAT CCG GTC CCC
     K   A   A   R   A   V   G   G   A   M   R   S   N   P   V   P
         130                                         140
448 ATC CTC ATC CCC TGC CAC AGG GTG GTT CGC AGT GAC GGT GCC ATC GGC
     I   L   I   P   C   H   R   V   V   R   S   D   G   A   I   G
                 150                                         160
496 CAT TAC ACC GGA GGA GGG CAG GCT GTG AAG GAG TGG CTT CTG GCC CAT
     H   Y   S   G   G   G   Q   A   V   K   E   W   L   L   A   H
                                     170
544 GAG GGC ATC CCG ACC GGA CAG CCA GCC TCC AAG GGC TTG GGT CTG ACT
     E   G   I   P   T   G   Q   P   A   S   K   G   L   G   L   T
             180                                         190
592 GGG ACC TGG CTC AAG TCA TCC TTC GAG TCG ACC AGC TCT GAG CCG TCT
     G   T   W   L   K   S   S   F   E   S   T   S   S   E   P   S
                         200
640 GGC CGA AAT TGA GTAACCGTTTGAATGACACATAGACGTAACGCGTGTCGGAAGCGGAT
     G   R   N   *
         210
699 GTGTGGTGGCACCACTATATTAAAAGAGCTGCAAGTGTCCTGGGGGACGCGACGTGGCTGCCC

762 TTTCTTGTTTCTACATTTTACAGCAGAATGAGTTCAGACGACCCACCTCGAGTGTGTCATCAG

825 TTCCCTCCTCTGACATGGTTATTGTTCTATTTTTAATGCCCATTAAAATGGGGGTGAGCGAGG

888 AAGGCCTGATGAGCCTTAGGGCCCCTTTGTATGAAGGGAGACACATCTGACACCTGCAG
```

FIGURE 2: Nucleotide sequence of the cDNA clone of mouse MGMT with the deduced amino acid sequence (by single-letter codes).

```
MOUSE    1 MAETCYMKYSVLDSPLGKMELSGCERGLHGIRLLSGKTPNTDPTEAPATP  50
           |   | ||   |||||||:|||||| |||:|:||:  |   |: | || :
HUMAN    1 MDKDCEMKRTTLDSPLGKLELSGCEQGLHEIKLLGKGTSAADAVEVPAPA  50

        51 EVLGGPEGVPEPLVQCTAWLEAYFREPAATEGLPLPALHHPVFQQDSFTR 100
            ||||    ||||:|||||||:|||::| | |::|:||||||||||:|||| 
        51 AVLGG....PEPLMQCTAWLNAYFHQPEAIEEFPVPALHHPVFQQESFTR  96

       101 QVLWKLLKVVKFGETVSYQQLAALAGNPKAARAVGGAMRSNPVPILIPCH 150
           |||||||||||||| :|||||||||||||||||||||||||:|||||||||| 
        97 QVLWKLLKVVKFGEVISYQQLAALAGNPKAARAVGGAMRGNPVPILIPCH 146

       151 RVVRSDGAIGHYSGGGQAVKEWLLAHEGIPTGQPA.SKGLGLTGTWLKSS 199
           ||| | ||:|:|||| |||||||||||   | |: : : || | |||:
       147 RVVCSSGAVGNYSGGL.AVKEWLLAHEGHRLGKPGLGGSSGLAGAWLKGA 195

       200 FESTSSEPSGRN 211
           :| | |||
       196 GATSGSPPAGRN 207
```

FIGURE 3: Comparison of the deduced amino acid sequence of mouse MGMT with that of human MGMT (Tano et al., 1990). The identical amino acids are marked by vertical lines, and similar ones are marked by colons.

near-homogeneity from BL21(DE3) possessing high-expression plasmid and found to have a size of about 24 kDa on SDS-PAGE, a value in good agreement with the molecular mass calculated from the predicted amino acid sequence.

The amino acid sequence of the N-terminal region was determined as described by von Wronski et al. (1991). The sequence determined, NH_2-Ala-Glu-Thr-X-Lys-Met-Lys-Tyr-Ser-Val-Leu-Asp-Ser-Pro-Leu-Gly-Lys-Met-Glu-Leu-Ser-Gly-X-Glu, uniquely placed an alanine codon at position 19. The initial methionine residue of the predicted sequence was not detected in the protein and was presumably removed by posttranslational cleavage by the methionine-specific aminopeptidase (Miller et al., 1987).

A comparison of amino acid sequences of mouse and human MGMTs shows an extensive homology along the entire length of the polypeptide except for an additional tetrapeptide seg-

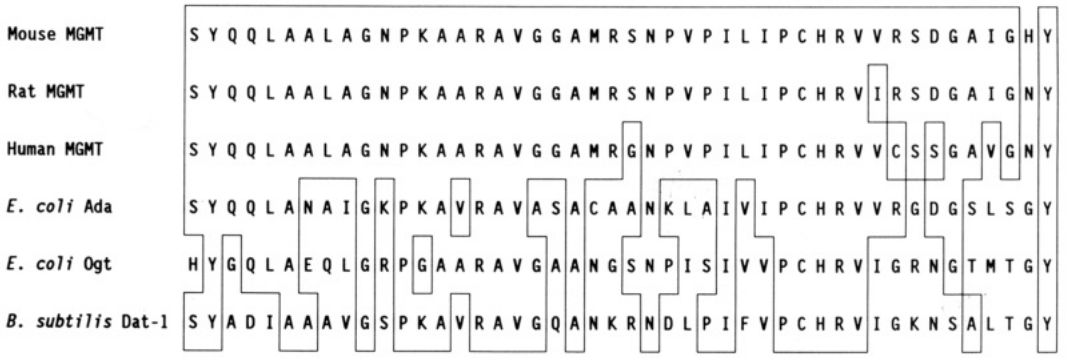

FIGURE 4: Sequence similarity of mammalian and bacterial methyltransferase. Amino acid residues 117–162 of mouse MGMT, 117–162 of rat MGMT (Potter et al., 1991), 113–158 of human MGMT, 289–334 of *E. coli* Ada protein (Demple et al., 1985), 107–152 of Ogt protein (Potter et al., 1987), and 98–143 of *B. subtilis* Dat-1 protein (Morohoshi et al., 1989) were aligned for maximum fit. Amino acids identical to mouse MGMT protein are shown boxed.

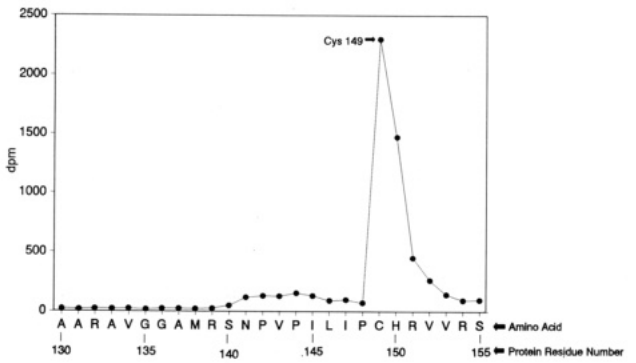

FIGURE 5: Identification of the alkyl-acceptor residue. Radioactivity in 50% aliquots of samples from each cycle of sequencing of the 4-kDa peptide was measured in a sintillation spectrometer and is plotted versus the amino acid identified in that cycle. The residue numbers are based on the predicted sequence given in Figure 2.

ment in the mouse protein. As indicated in Figure 3, the tetrapeptide segment has the sequence Pro-Glu-Gly-Val. However, because of the Pro-Glu sequence repeat, the human and mouse protein sequences could also be aligned such that the additional peptide in the mouse protein has the sequence Gly-Val-Pro-Glu. Of the 207 amino acid pairs in the alignment, 70% (i.e., 145) were accounted for by identical residues. In particular, in the region between amino acids 117 and 162, the mouse protein exhibits more than 89% homology with the human MGMT and is identical to rat MGMT except for two residues. Furthermore, the mouse protein shows a significant homology to that of the bacterial MGMTs (Figure 4). Among the 46 residues around the alkyl acceptor site, 27 residues of the C-terminal half of Ada protein, 24 residues of the Ogt protein, and 23 residues of the Dat-1 protein are identical to the residues of mouse protein. Five amino acid residues including the alkyl acceptor cysteine are completely conserved in all six proteins. It is thus predicted from these sequence similarities that the Cys-149 of mouse MGMT protein is the acceptor residue for the alkyl group from O^6-alkylguanine in alkylated DNA.

Identification of the Alkyl Acceptor Cysteine Residue in Mouse MGMT. Cysteine has been found to be the only alkyl acceptor residue in all MGMTs (Lindahl et al., 1988). The amino acid sequence deduced from the cDNA sequence indicates the presence of four cysteine residues at positions 5, 24, 66, and 149 in the mouse MGMT. Analysis of lysyl endopeptidase cleaved peptides from [CH_3-^{3}H]-methylated MGMT by SDS–polyacrylamide gel electrophoresis and autoradiography showed the presence of only one 4-kDa radioactive band (data not shown). A 4-kDa peptide fragment with

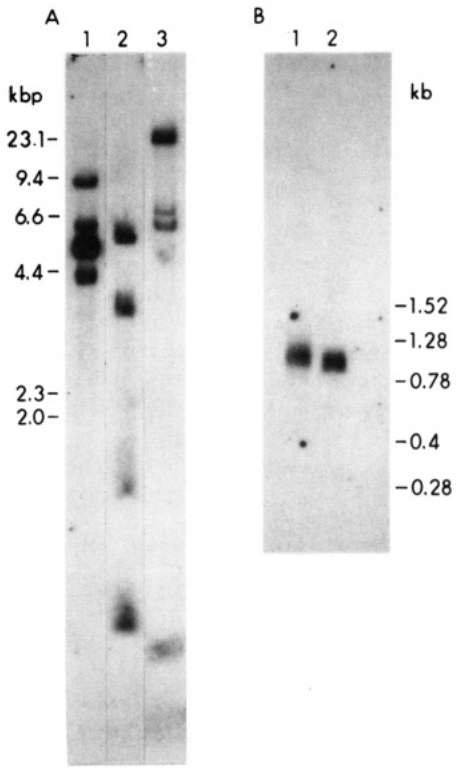

FIGURE 6: Southern and Northern blot analysis with pSS601. Hybridization was carried out as described under Materials and Methods. (A) Southern analysis of mouse genomic DNA. Ten micrograms of DNA derived from the spleen of Balb/C digested with *Eco*RI (lane 1), *Pst*I (lane 1), and *Hind*III (lane 3). (B) Northern blot analysis of total RNA derived from mouse NIH3T3 (lane 1) and HeLaS3 cells (lane 2). Twenty micrograms of each RNA was used.

40 residues (Ala-130 through Val-169) was predicted from the amino acid sequence of the mouse protein and includes Cys-149, the candidate alkyl acceptor residue. Further sequence analysis of the peptide and measurement of radioactivity of separated amino acid derivatives confirmed the amino acid sequence predicted from the cDNA sequence and indicated that Cys-149 was the only radioactive residue (Figure 5).

The absence of radioactivity in the 7-kDa peptide that should include Cys-66 eliminates the latter as an alkyl acceptor residue. Analysis of the N-terminal sequence of the intact methylated protein showed that Cys-5 and Cys-24 were also not radiolabeled (data not shown). These data thus established unequivocally that Cys-149 is the only alkyl acceptor residue in mouse MGMT.

Table I: MGMT Activities Encoded by Plasmids Containing Wild-Type and Mutant cDNA in *E. coli* Extracts

E. coli host	vector	plasmid	IPTG[a] treatment	sp act. (pmol/mg of protein)
JM109	pUC18		−	1
JM109	pUC18		+	1
JM109	pUC18	pSS601	−	87
JM109	pUC18	pSS601	+	734
JM109	pUC18	pSS6011	−	53
JM109	pUC18	pSS6011	+	536
BL21(DE3)	pET11a	pSS700	−	5
BL21(DE3)	pET11a	pSS700	+	3092
BL21(DE3)	pET11a	pSS701	−	18
BL21(DE3)	pET11a	pSS701	+	1820

[a] IPTG, isopropyl β-D-thiogalactoside.

Hybridization Analysis of Mouse Gene Sequences and Transcript. Southern blot analysis of the genomic DNA from a Balb/C mouse probed with mouse *Mgmt* cDNA showed the presence of both strongly and weakly hybridizing *Eco*RI, *Hin*dIII, and *Pst*I fragments (Figure 6A). The genomic DNAs from two other mouse strains (C3Hf and C57BL/6) also showed the identical hybridization pattern (data not shown). Northern blot analysis of total RNAs from mouse NIH3T3 cells showed the presence of a single 1.0-kb-long transcript which was slightly larger than that of human MGMT (Figure 6B).

Construction of a High-Level Expression Plasmid for Mouse MGMT. In order to purify large amounts of mouse MGMT protein, we constructed a high-level expression plasmid in a vector, pET11a, designed by Studier et al. (1990). The procedure involved oligonucleotide-mediated site-directed mutagenesis such that a new *Nde*I recognition sequence, CATATG, was created at the translation start site. The PCR protocol of Hemsley et al. (1989) was used to create the mutation. The pSS601 was mutagenized and amplified with PCR using two primers (see Materials and Methods). After PCR reaction, the ends of PCR products were filled with *E. coli* Klenow fragment and four dNTPs and then precipitated with ethanol. One-fourth of the end-filling PCR products were circularized with T4 DNA ligase and were used for transformation of *E. coli* DH5α cells. The plasmid having the new *NdeI* recognition site, pSS6011, has the same level of MGMT activity as its parental plasmid, pSS601, in JM109.

A 900 bp *Nde*I fragment of pSS6011, containing about 200 bp of vector segment downstream of the *Mgmt* cDNA, was inserted into the *Nde*I site of pETlla to make a high-level expression plasmid for mouse MGMT. The resulting plasmid, pSS700, having the correct orientation with respect to the T7 promoter was used for purification of protein.

Activity of MGMT Encoded by cDNA. Table I shows the mouse MGMT activity in *E. coli* crude extracts encoded by pSS601, pSS6011, pSS700, and the control plasmids. IPTG induction of the mouse MGMT encoded by pSS601 indicates that MGMT expression from this plasmid is controlled by the *lac* promoter. A pSS6011 containing the new *Nde*I recognition sequence has the same level of MGMT activity as its parental plasmid pSS601.

After IPTG treatment of *E. coli* cells harboring high-expression plasmid, the level of MGMT increased 4–5 times higher than that of pSS601 or pSS6011. Induced MGMT protein is easily detected by SDS-PAGE analysis of crude extract (data not shown). The expression of pSS700 in BL21(DE3) without IPTG is highly suppressed in comparison with pSS601 and pSS6011 in JM109.

Methyltransferase Activity of Mutated Mouse MGMT. To investigate the role of the unique tetrapeptide sequence Pro-Glu-Gly-Val (or Gly-Val-Pro-Glu) present in mouse MGMT, we constructed a mutant of the mouse *Mgmt* cDNA by oligonucleotide-mediated site-directed mutagenesis by the method of inverse PCR (Hemsley et al., 1989) using primers 5′GAGGTGCTCGGTGGGCCAGAGCCTC TGGTG3′ and 5′AGGAGTAGCTGGGGCCTCTGT3′. The longer primer has the same sequence as that of positions 161–207 of the sense strand except for deletion of 12 nucleotides (5′CCAGAGGGAGTT3′) corresponding to the tetrapeptide. Another primer is exactly the same as that of the opposite strand starting with the adjacent 5′ end of the longer one as described under Materials and Methods. The PCR reaction and the subsequent procedures were also described earlier. The mutated plasmid, pSS701, was cloned, and the deletion of the 12 bp segment corresponding to the PEGV sequence was confirmed by nucleotide sequencing of the cDNA insert. There was no significant difference in the MGMT activities of the wild-type and mutant MGMT proteins (Table I). Amplification of the mouse cDNA by PCR in an independent experiment led to an adventitious mutation of codon 169 from GTG to GCG. The corresponding change from valine to alanine in the MGMT polypeptide did not affect the activity of the mutant protein expressed in *E. coli* (data not shown).

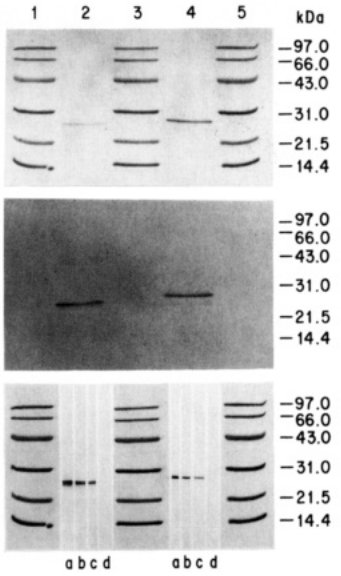

FIGURE 7: SDS–PAGE of purified recombinant human (lane 2) and mouse (lane 4) MGMT. Lanes 1, 3, and 5 contained Bio-Rad molecular mass markers. Proteins (4 μg of human, 1.3 μg of mouse) were separated in a 10% gel by the method of Laemmli and electroblotted onto a PVDF membrane. The *top panel* shows the proteins stained with Coomassie Blue R-250. The *middle panel* is the corresponding autoradiograph. The *bottom panel* shows the results of immunoprobing. Lanes 2 and 4 were cut into four strips, completely destained with methanol, and probed with the following mAbs: (a) 4.A1; (b) 19.2; (c) 21.A8; (d) control IgG. Antibody binding was visualized by a silver-enhanced gold method as described under Materials and Methods.

Antigenic Cross-Reactivity of Mouse and Human MGMTs. Purified recombinant mouse MGMT protein was examined for cross-reactivity with monoclonal antibodies specific for human MGMT protein. Autoradiography of a Western blot confirmed that the mouse cDNA-encoded protein was the methyl acceptor protein, and was slightly larger than human MGMT (Figure 7, middle panel). Lanes 2 and 4 containing human and mouse MGMT, respectively, were cut into four strips and immunoprobed with three different monoclonal antibodies specific for human cellular MGMT and with nonimmune IgG. The bottom panel of Figure 7 indicated that all three monoclonal antibodies reacted with the mouse

MGMT expressed in *E. coli* from the cloned cDNA with affinity comparable to that for human recombinant MGMT.

Discussion

The cloning of a mouse cDNA encoding the DNA repair protein MGMT responsible for removal of mutagenic and carcinogenic O^6-alkylguanine is essential not only for comparing the structure and regulation of human and mouse MGMT genes but also for experimental manipulation of regulation of the mouse gene in cultured cells and in whole animals. Furthermore, large-scale production of wild-type and mutant mouse protein in *E. coli* from appropriate cDNA expression vectors should provide an opportunity for comparative biochemical studies which in turn may lead to an understanding of the structure–function relationship of these unusual proteins. The 1.5-kb mouse *Mgmt* cDNA cloned from a cDNA library was larger than the 1.0-kb MGMT mRNA (Figure 6B). It also lacked the 3′-terminal sequence of the mRNA including the poly(A) sequence and was apparently fused to an unrelated cDNA sequence. Some of the 5′-terminal sequence may also be missing in the cDNA. However, investigation of the open reading frame indicated three possible translation start sites within the cDNA sequence. With the ATG codon at nucleotide 16, a polypeptide of 211 residues was predicted while significantly shorter polypeptides were predicted from the ATG codon of positions 34 and 70. N-Terminal sequence analysis of the cDNA-encoded polypeptide expressed in *E. coli* established that translation was initiated from the ATG codon at position 16. The mouse protein of 22 kDa is larger than the human MGMT by 0.4 kDa, and the slight difference in size can be seen during SDS–PAGE analysis. It is interesting that the mouse MGMT expressed in *E. coli* with a penultimate alanine residue in the predicted amino-terminal sequence did not contain the terminal methionine residue, unlike the human protein which has aspartic acid as the penultimate amino-terminal residue (Tano et al., 1990; von Wronski et al., 1991). It appears that the methionine residue in the mouse protein was cleaved posttranslationally by an aminopeptidase of *E. coli* that removes amino-terminal methionine when the second residue is alanine, threonine, or glycine (Miller et al., 1987). The mouse MGMT polypeptide has 70% sequence homology with the human protein and 91% homology with the rat protein (Potter et al., 1991). In contrast to this homology among the mammalian proteins throughout the polypeptides, the sequence homology with the bacterial MGMTs is mostly limited to the region which spans the alkyl-acceptor cysteine residue in the case of the bacterial and human MGMTs. In particular, the sequence PCHRV is conserved in all MGMTs which include the alkyl-acceptor cysteine. Thus, it appeared more than likely that Cys-149 of the mouse MGMT is the alkyl-acceptor residue. We confirmed this possibility by direct sequence analysis of the only radioactive peptide in the methylated protein (Figure 5). We also eliminated the possibility that the other three cysteine residues present in the protein may also accept methyl groups. A additional benefit of the experiment was the confirmation of the sequence of 26 residues surrounding the alkyl acceptor site that was predicted from the cDNA sequence. The role of the other residues in the conserved pentapeptide sequence in MGMT activity can be tested by site-directed mutagenesis of the cloned plasmid. We have already shown that in the case of *E. coli* Ada protein, the cysteine residue could not be substituted with another nucleophilic amino acid histidine and that the inversion of the cysteine-histidine dipeptide sequence also resulted in an inactive protein (Tano et al., 1989).

The sequence similarity among the mammalian proteins predicts a high degree of antigenic cross-reactivity. Indeed, three monoclonal antibodies raised against human MGMT recognize the mouse MGMT in immunoblots, indicating the presence of common epitopes in the two proteins. However, the sequences of these proteins are significantly divergent in the C-terminal region. Thus, it is not surprising that a rabbit polyclonal antibody against a synthetic peptide based on the human MGMT sequence (residues 171–184) did not cross-react with the mouse protein (Ostrowski et al., 1991). Only 5 of 14 of the amino acid residues in this peptide sequence were identical in the 2 proteins.

We have constructed a high-expression vector for large-scale production of native MGMT for physicochemical studies of the protein. While PCR provides a very powerful approach for site-directed mutagenesis, the technique also has a high risk of adventitious mutations induced during DNA amplification and cloning of amplified DNA. We thus obtained an unexpected mutation of Val-169 to Ala-169. Val-169 is a conserved residue in rat and human MGMTs as well. The fact that the Ala-169 mutant retains methyltransferase activity suggests that the hydrophobic side chain of Val-169 is not directly involved at the active site.

The mouse protein expressed in *E. coli* and containing 210 amino acid residues (lacking the amino-terminal methionine residue) is larger than the human MGMT by 3 residues and the rat MGMT by 1 or 2 residues. It is interesting that in a region of highly conserved sequence (residues 51–63 in the mouse MGMT) the rodent proteins have a sequence of PEGVPE compared to only PE in the human protein. Thus, the additional tetrapeptide in rodent proteins could be either PEGV or GVPE, and deletion of either one leads to the same mutant protein. In the present study, we generated a mutant mouse MGMT in which the PEGV sequence was deleted. Our preliminary results show that the deletion mutant retained MGMT activity. However, we have not carried out detailed analysis to see whether the kinetic properties of the protein were altered as a result of the mutation.

Acknowledgments

We thank Dr. R. S. Foote for oligonucleotide synthesis. We also thank W. C. Dunn for excellent technical assistance.

References

Bhattacharyya, D., Foote, R. S., Boulden, A. M., & Mitra, S. (1990) *Eur. J. Biochem. 193*, 337–343.

Bogden, J. M., Eastman, A., & Bresnick, E. (1981) *Nucleic Acids Res. 9*, 3089–3103.

Brent, T. P. (1985) *Pharmacol. Ther. 31*, 121–140.

Brent, T. P., Houghton, P. J., & Houghton, J. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2985–2989.

Brent, T. P., von Wronski, M. A., Pegram, C. N., Bigner, D. D. (1990) *Cancer Res. 50*, 58–61.

Chomczynski, P., & Sacchi, N. (1987) *Anal. Biochem. 162*, 156–159.

D'Ambrosio, S. M., Samuel, M. J., Dutta-Choudhury, T. A., & Wani, A. A. (1987) *Cancer Res. 47*, 51–57.

Day, R. S., Ziolkowski, C. H. J., Scudiero, D. A., Meyer, S. A., Lubiniecki, A. S., Girardi, A. J., Galloway, S. M., & Bynum, G. D. (1980) *Nature 288*, 724–727.

Demple, B., Sedgwick, B., Robins, P., Totty, N., Waterfield, M. D., & Lindahl, T. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2688–2692.

Feinberg, A. P., & Vogelstein, B. (1983) *Anal. Biochem. 132*, 6–13.

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.
Green, E. L. (1981) in *Genetics and probability in animal breeding experiments*, pp 77–113, Macmillan, New York.
Hayakawa, H., Koike, G., & Sekiguchi, M. (1990) *J. Mol. Biol. 213*, 739–747.
Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G., & Galas, D. J. (1989) *Nucleic Acids Res. 17*, 6545–6551.
Kodama, K., Nakabeppu, Y., & Sekiguchi, M. (1989) *Mutat. Res. 218*, 153–163.
Laemmli, U. K. (1970) *Nature 227*, 680–685.
Lindahl, T., Sedgwick, B., Sekiguchi, M., & Nakabeppu, Y. (1988) *Annu. Rev. Biochem. 57*, 133–157.
Loveless, A. (1969) *Nature (London) 233*, 206–207.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning, a Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Margison, G. P., Cooper, D. P., & Brennand, J. (1985) *Nucleic Acids Res. 13*, 1939–1952.
Matsudaira, P. (1987) *J. Biol. Chem. 262*, 10035–10038.
Miller, C. G., Strauch, K. L., Kukral, A. M., Miller, J. L., Wingfield, P. T., Mazzei, G. J., Werlen, R. C., Graber, P., & Movva, N. R. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 2718–2722.
Mitra, G., Pauly, G. T., Kumar, R., Pei, G. K., Hughes, S. H., Moschel, R. C., & Barbacid, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 8650–8654.
Morohoshi, F., Hayashi, K. & Munakata, N. (1989) *Nucleic Acids Res. 17*, 6531–6543.
Nakabeppu, Y., & Sekiguchi, M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 6297–6301.
Olsson, M., & Lindahl, T. (1980) *J. Biol. Chem. 255*, 10569–10571.
Ostrowski, L. E., Pegram, C. N., von Wronski, M. A., Humphrey, P. A., He, X., Shiota, S., Mitra, S., Brent, T. P. & Bigner, D. D. (1991) *Cancer Res. 51*, 3339–3344.
Pegg, A. E. (1990) *Cancer Res. 50*, 6119–6129.
Pegg, A. E., & Singer, B. (1984) *Cancer Invest. 2*, 223–236.
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.
Potter, P. M., Rafferty, J. A., Cawkwell, L., Wilkinson, M. C., Copper, D. P., O'Connor, P. J., & Margison, G. (1991) *Carcinogenesis 12*, 727–733.
Rebeck, G. W., Coons, S., Carroll, P., & Samson, L. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 3039–3043.
Rydberg, B., Spurr, N., Karran, P. (1990) *J. Biol. Chem. 265*, 9563–9569.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Schagger, H., & von Jagow, G. (1987) *Anal. Biochem. 166*, 368–379.
Schold, S. C., Jr., Brent, T. P., von Hofe, E., Friedman, H. S., Mitra, S., Bigner, D. D., Swenberg, J. A., & Kleihues, P. (1989) *J. Neurosurg. 70*, 573–577.
Sklar, R., & Strauss, B. (1981) *Nature 289*, 417–420.
Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60–89.
Tano, K., Foote, R. S., & Mitra, S. (1988) *Gene 64*, 305–311.
Tano, K., Bhattacharyya, D., Foote, R. S., Mural, R. J., & Mitra, S. (1989) *J. Bacteriol. 171*, 1535–1543.
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., Bigner, D. D., & Brent, T. P. (1989) *Cancer Commun. 1*, 323–327.
von Wronski, M. A., Shiota, S., Tano, K., Mitra, S., Bigner, D. D., & Brent, T. P. (1991) *J. Biol. Chem. 266*, 1064–1070.
Washington, W. J., Foote, R. S., Dunn, W. C., Generoso, W. M., & Mitra, S. (1989) *Mech. Ageing Dev. 48*, 43–52.