Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 9127–9131, October 1991
Genetics

Cloning and characterization of a 3-methyladenine DNA glycosylase cDNA from human cells whose gene maps to chromosome 16

(human DNA alkylation repair gene)

LEONA SAMSON*, BRUCE DERFLER, MICHAEL BOOSALIS, AND KATHERINE CALL

Molecular and Cellular Toxicology, Harvard School of Public Health, 665 Huntington Avenue, Boston, MA 02115

Communicated by Elkan R. Blout, July 1, 1991 (received for review April 25, 1991)

ABSTRACT **We described previously the isolation of a *Saccharomyces cerevisiae* 3-methyladenine (3-MeAde) DNA glycosylase repair gene (*MAG*) by its expression in glycosylase-deficient *Escherichia coli alkA tag* mutant cells and its ability to rescue these cells from the toxic effects of alkylating agents. Here we extend this cross-species functional complementation approach to the isolation of a full-length human 3-MeAde DNA glycosylase cDNA that rescues *alkA tag E. coli* from killing by methyl methanesulfonate, and we have mapped the gene to human chromosome 16. The cloned cDNA, expressed from the pBR322 β-lactamase promoter, contains an 894-base-pair open reading frame encoding a 32,894-Da protein able to release 3-MeAde, but not 7-methylguanine, from alkylated DNA. Surprisingly, the predicted human protein does not share significant amino acid sequence homology with the bacterial AlkA and Tag glycosylases or the yeast MAG glycosylase, but it does share extensive amino acid sequence homology with a rat 3-MeAde DNA glycosylase and significant DNA sequence homology with genes from several mammalian species. The cloning of a human 3-MeAde DNA glycosylase cDNA represents a key step in generating 3-MeAde repair-deficient cells and the determination of the *in vivo* role of this DNA repair enzyme in protecting against the toxic and carcinogenic effects of alkylating agents.**

The genome of every organism continually sustains DNA damage which, if left unrepaired, contributes to cell death, mutation, chromosome damage, ageing, and carcinogenesis. Thus, a number of DNA repair pathways have evolved, and these appear to be highly conserved among bacteria, yeast, insect, and mammalian cells. One of the pathways for DNA repair is initiated by the action of certain glycosylases that excise abnormal bases from DNA, leaving behind apurinic or apyrimidinic sites that then trigger nucleotide excision repair (1). Abnormal DNA bases are known to be continually produced by uracil misincorporation, spontaneous bond breakages, and reactions with normal cellular metabolites and environmental DNA-damaging agents (2). At least eight types of glycosylase have been identified, each of which is specific for the removal of one or more abnormal bases (2). 3-Methyladenine (3-MeAde) is one of the major lethal lesions produced by agents like methyl methanesulfonate (MMS) and *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG) in *Escherichia coli*, and this lesion is also believed to be produced by the normal cellular metabolite *S*-adenosylmethionine; *E. coli* is protected against such alkylating agents by the constitutively expressed *tag* gene and the alkylation-inducible *alkA* gene, which encode two 3-MeAde DNA glycosylases (3). Whether the 3-MeAde lesion is lethal or mutagenic in human cells is not yet known because mutant human cells deficient in 3-MeAde DNA glycosylase have not yet been identified. Further, it remains to be determined whether the 3-MeAde DNA glycosylase provides any protection against the carcinogenic effects of alkylating agents in our environment. Here we describe a key step towards defining the *in vivo* role of the human 3-MeAde DNA glycosylase repair enzyme—namely, the cloning of a cDNA coding for this DNA repair function.

A number of human DNA repair genes have been cloned by their ability to rescue DNA damage-sensitive rodent cells from the toxic effects of the appropriate DNA-damaging agent (4, 5). However, this approach is limited by the fact that the specific DNA repair defects in the rodent cell lines are unknown (5), and this has made the *in vivo* role of the cloned gene products difficult to determine. We recently developed a method for cloning eukaryotic DNA repair genes that detects expression of the cloned genes by their ability to rescue well-characterized DNA repair-deficient strains of *E. coli* from the toxic effects of DNA damage (6). Using this procedure, we cloned a 3-MeAde DNA glycosylase gene (*MAG*) from *Saccharomyces cerevisiae* by its ability to rescue a 3-MeAde DNA glycosylase-deficient *E. coli alkA tag* double mutant from the killing effects of MMS. The *S. cerevisiae* MAG glycosylase shares significant sequence homology with the *E. coli* AlkA glycosylase, and its expression, like the transcription of *alkA*, is dramatically increased when the yeast is exposed to alkylating agents (7). To determine the *in vivo* role of 3-MeAde repair, we used the cloned *MAG* gene to produce *mag*$^-$ yeast cells and showed that the reduction in 3-MeAde repair resulted in increased sensitivity to alkylation-induced cell death but not mutation in *S. cerevisiae* (6, 7). We also found that *mag*$^-$ cells express another 3-MeAde DNA glycosylase (7), but it remains to be determined whether this second glycosylase represents a Tag homologue.

We have now extended this eukaryote/prokaryote functional complementation approach to the isolation of a human 3-MeAde DNA glycosylase cDNA. It was not previously possible to clone this human gene by conventional methods because 3-MeAde DNA glycosylase-deficient mammalian cell lines are not currently available. Here we describe the characterization of a full-length human 3-MeAde DNA glycosylase cDNA† isolated by its ability to rescue *E. coli alkA tag* mutants from killing by MMS, and we have mapped its gene to human chromosome 16. The predicted human 3-MeAde DNA glycosylase enzyme shares no amino acid homology with the *E. coli* AlkA and Tag glycosylases or the *S. cerevisiae* MAG glycosylase, but it does share significant homology with the rat 3-MeAde DNA glycosylase enzyme (8).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: 3-MeAde, 3-methyladenine; MMS, methyl methanesulfonate; MNNG, *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine; 7-MeGua, 7-methylguanine.

*To whom reprint requests should be addressed.

†The sequence reported in this paper has been deposited in the GenBank data base (accession no. M74905).

MATERIALS AND METHODS

Cells, Plasmids, and Enzymes. *E. coli* AB1157 (F⁻ *thr-1 leu-6 proA2 thi-1 argE lacY1 galK ara-14 xyl-5 mly-1 tsx-33 strA sup-37*) was the strain used as the wild type for alkylation. MV1932 and MV1902 (gifts from M. Volkert, University of Massachusetts, Worcester, MA) are alkylation-sensitive derivatives of AB1157 and are *alkA1 tag* and *alkA105*::λpSG1 (9), respectively. The human cDNA expression library was constructed by Prochownik *et al.* (10) by cloning human liver cDNAs under the β-lactamase promoter of the pBR322 derivative pKT218 (11) and was a gift from Stuart Orkin (Harvard Medical School, Boston). Restriction endonucleases were from New England Biolabs, and digestions were carried out by standard protocols (12).

Screen for Alkylation-Resistant Transformants. Two hundred nanograms of the cDNA library was transformed into *E. coli alkA tag* (MV1932); at least 10^7 independent transformants were incubated for 3.5 hr at 37°C to allow cDNA insert expression and then were plated on six large Luria-Bertoni (LB) ampicillin plates containing 0.005–0.01% MMS. Approximately 8000 surviving colonies were scraped from the plates, plasmid DNA (pool 1) was isolated and reintroduced into MV1932, and the process was repeated to generate DNA pool 2. Pool 2 DNA was reintroduced into MV1932, and transformants were plated on LB ampicillin plates without MMS; 300 transformants were individually screened for MMS resistance as described (6), and several transformants were found to be MMS resistant. Each plasmid in the resistant transformants contained the same cDNA insert (data not shown). One plasmid, called pP5-3, was used for detailed analysis.

Survival Curves and DNA Glycosylase Activity. For survival measurements, bacteria were grown in LB medium to 10^8 cells per ml; MMS was added to 0.05%, and aliquots were removed at various times, diluted in M9 salts, and spread on LB plates to estimate cell survival. For measurements of 3-MeAde DNA glycosylase activity, cell extracts were prepared as described (13) in 50 mM Hepes adjusted with KOH to pH 7.6/100 mM KCl/1 mM EDTA/5 mM dithiothreitol. Extract protein was incubated for 1 hr at 37°C with 33,000 cpm of di[^{3}H]methyl sulfate-treated calf thymus DNA (250 cpm/μg) prepared according to Samson and Linn (14), and the release of 3-MeAde and 7-methylguanine (7-MeGua) was measured by paper chromatography as described (6).

Southern and Northern Blot Analysis. Bacterial (15), mammalian (12), and yeast (16) DNAs were isolated as described, and the corresponding RNAs were isolated as described (12). A "zoo blot" purchased from Clontech contained the *Eco*RI-digested human, monkey, rat, mouse, dog, cow, rabbit, chicken, and yeast DNAs (8 μg per lane). A mapping panel blot, purchased from Bios (New Haven, CT), consisted of *Bam*HI-digested DNA of human, hamster, and 25 human–hamster hybrids. DNA from human–mouse hybrid GM10567 (containing human chromosome 16) was purchased from the NIGMS Human Genetic Cell Repository, Camden, NJ. Northern blot analysis was as described (12). The blots were probed with a ^{32}P-labeled 0.5-kilobase (kb) *Pst* I fragment from the pP5-3 cDNA insert, and the final filter washes were at high stringency.

DNA Sequence Analysis. Three *Pst* I fragments of 0.3, 0.4, and 0.5 kb were subcloned from the pP5-3 cDNA insert into pTZ18R and pTZ19R (Pharmacia, LKB), transformed into *E. coli* NM522, and sequenced in both directions. We confirmed the fragment order by sequencing across the *Pst* I sites in pP5-3, using the appropriate oligonucleotide primers to sequence from the double-stranded pP5-3 plasmid.

RESULTS AND DISCUSSION

Isolation of a Human 3-MeAde DNA Glycosylase cDNA. Our aim was to isolate a human cDNA whose expression rescues 3-MeAde-DNA-glycosylase-deficient *alkA tag E. coli* from killing by MMS. The expression library was constructed by Prochownik *et al.* (10) by cloning human liver cDNAs under the β-lactamase promoter of the pBR322 derivative pKT218 (11); each cDNA insert replaced most of the β-lactamase coding sequence and could potentially be expressed either as a fusion protein or from its own ATG translation initiation codon. The library was transformed into the *E. coli alkA tag* strain MV1932 (9), and a population of 10^7 independent transformants was enriched for alkylation-resistant cells by repeated challenges with MMS. Individual colonies from the enriched population were streaked onto MMS plates and control plates; several transformants formed colonies on the MMS plates, and the responsible plasmids were isolated from the control plate bacteria and checked for their ability to transmit MMS resistance. In this way we identified plasmid pP5-3 with a 1.2-kb cDNA insert that rescued *alkA tag E. coli* from the killing effects of MMS (Fig. 1*A*) and MNNG (data not shown). Although *alkA tag*/pP5-3 cells were considerably more resistant than *alkA tag* cells, they were not as resistant as wild-type AB1157 cells. We eliminated the pos-

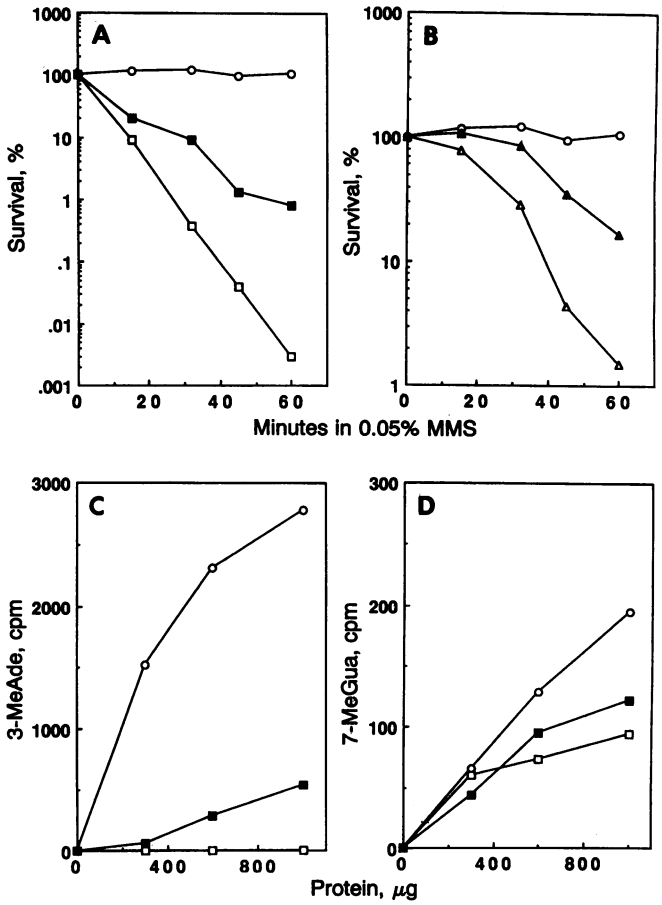

FIG. 1. Characterization of the phenotype conferred by pP5-3. (*A* and *B*) Bacteria were grown in LB broth to 10^8 cells per ml, MMS was added to 0.05%, and aliquots were removed at various times, diluted, and plated on LB agar to estimate cell survival. The strains are: ○, AB1157 (wild type); □, MV1932 (*alkA tag*); ■, MV1932 (*alkA tag*/pP5-3); △, MV1902 (*alkA105*::λpSG1); and ▲, MV1902 (*alkA105*::λpSG1/pP5-3). (*C* and *D*) Extract protein was incubated for 1 hr at 37°C with di[^{3}H]methyl sulfate-treated calf thymus DNA as described, and the release of 3-MeAde (*C*) and 7-MeGua (*D*) was measured by paper chromatography. Extracts were from AB1157 (○), MV1932 (*alkA tag*) (□), and MV1932 (*alkA tag*/pP5-3) (■).

sibility that pP5-3 conferred resistance to MV1932 by suppression of the point mutation in *alkA* or *tag* by showing that pP5-3 also conferred alkylation resistance to a nonsuppressible *E. coli* strain bearing a λ phage insertion in the *alkA* gene (Fig. 1*B*). Further experiments showed that pP5-3 encodes a 3-MeAde DNA glycosylase. Fig. 1*C* shows that *alkA tag*/pP5-3 cells contained 3-MeAde DNA glycosylase activity, although not as much as in wild-type AB1157 cells, which agrees with the observation that *alkA tag*/pP5-3 are not as alkylation-resistant as AB1157 (Fig. 1*C*). The pP5-3-encoded glycosylase in crude *E. coli* cell extracts did not release 7-MeGua from alkylated DNA (Fig. 1*D*), and it remains to be determined whether it releases 3-methylguanine (3-MeGua) or O^2-methylpyrimidines as the *E. coli* AlkA glycosylase does (3). It was previously shown that the partially purified human 3-MeAde DNA glycosylase releases 7-MeGua and 3-MeGua from alkylated DNA, in addition to 3-MeAde, albeit at a slow rate (17, 18); although our human glycosylase preparations did not release any 7-MeGua in crude *E. coli* cell extracts, this activity may be detectable when the enzyme is purified.

Characterization of the Human 3-MeAde DNA Glycosylase cDNA. A pP5-3 cDNA insert fragment hybridized to human genomic DNA (Fig. 2*A*) and to a 1.2- to 1.3-kb human mRNA (Fig. 2*B*), indicating that we had cloned a virtually full-length human cDNA encoding 3-MeAde DNA glycosylase. Hybridization of the pP5-3 cDNA also detected a 1.2-kb mRNA in baboon, which was expressed in all tissues examined (data not shown). The abundance of the mRNA varied as follows: spleen > muscle, liver and heart > kidney and brain. 3-MeAde DNA glycosylase activity also varies in mouse tissues, with spleen expressing up to 10-fold more activity than brain tissue (19). O^6-methylguanine DNA methyltransferase is the only other alkylation-specific DNA repair enzyme known in mammalian tissues, but it does not appear to be coordinately expressed with the 3-MeAde DNA glycosylase (at least in mice) because, while spleen has the highest level of glycosylase activity relative to other tissues, it has one of the lowest levels of O^6-methylguanine DNA methyltransferase activity (19–21).

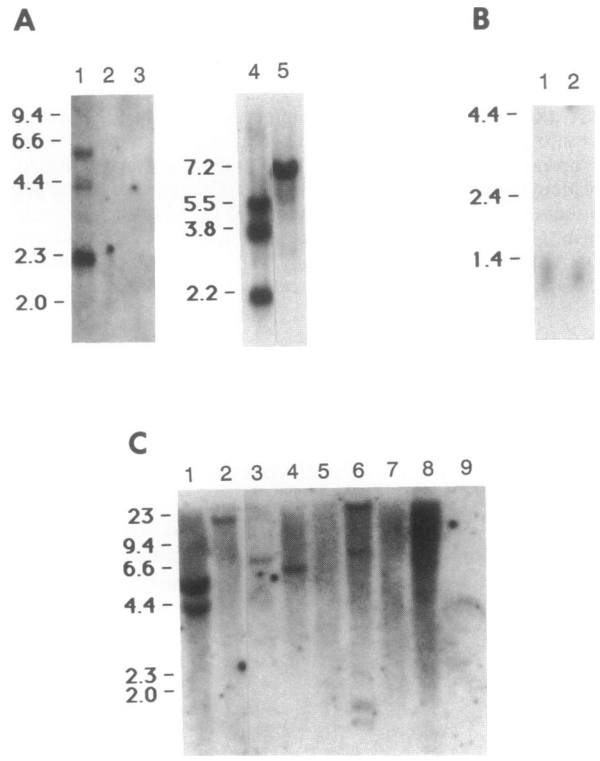

FIG. 2. Hybridization analysis of the pP5-3 cDNA insert. (*A*) Southern blot analysis. DNAs [10 μg of human (lanes 1 and 4), 10 μg of hamster (lane 5), 2.5 μg of *S. cerevisiae* (lane 2), and 5.0 μg of *E. coli* (lane 3)] were digested with *Bam*HI and probed with a ^{32}P-labeled 0.5-kb *Pst* I fragment from the pP5-3 cDNA insert. Final filter washes were at high stringency. (*B*) Northern blot analysis. RNA from HeLa CCL2 and HeLa S3 cells (30 μg of total RNA), separated in formaldehyde/1% agarose (12), was probed with a 0.5-kb *Pst* I fragment from the pP5-3 cDNA insert and washed at high stringency. (*C*) Southern zoo blot analysis. A "zoo blot", purchased from Clontech, was probed with a 0.5-kb *Pst* I fragment of the pP5-3 cDNA insert and was washed at high stringency. The *Eco*RI-digested DNAs (8 μg per lane) were from the following species: human (lane 1); monkey (lane 2); rat (lane 3); mouse (lane 4); dog (lane 5); cow (lane 6); rabbit (lane 7); chicken (lane 8); and yeast (lane 9).

The human glycosylase cDNA did not hybridize to *S. cerevisiae* or *E. coli* DNA (Fig. 2 *A* and *C*), but it did exhibit cross-species hybridization to six of eight vertebrate DNAs tested—namely, hamster (Fig. 2*A*), monkey, rat, mouse, cow, and chicken; surprisingly, it did not hybridize to dog or rabbit DNA (Fig. 2*C*). The cloned human cDNA detected the same size *Eco*RI band (7 kb) reported to hybridize to the rat 3-MeAde DNA glycosylase cDNA (8), suggesting that the DNA fragments hybridizing to the human glycosylase cDNA in the six positive species (Fig. 2*A*) probably represent homologous glycosylase genes.

The sequence of the cloned cDNA showed an 894-base-pair open reading frame, beginning with its own ATG and displaying a consensus AATAAA sequence to signal poly(A)$^+$ addition downstream of the TGA translation stop codon (Fig. 3*A*). The glycosylase cDNA open reading frame predicts a 32,894-Da enzyme. This is significantly larger than the 25-kDa enzyme isolated from human placenta (18), suggesting that there may be more than one human 3-MeAde DNA glycosylase or that the placental glycosylase is a cleavage product from a larger precursor. The amino acid sequence of the human glycosylase displayed no significant similarity to the *E. coli* AlkA or Tag glycosylases or the *S. cerevisiae* MAG glycosylase. However, a 176-amino acid stretch of the human glycosylase shares 85% identity with the middle of the predicted rat glycosylase polypeptide (Fig. 3*B*) (8). Clearly the rat and human glycosylases share a common origin, and we have named the human gene *AAG* for 3-alkyladenine DNA glycosylase. Further analysis showed that two frameshifts in the rat cDNA sequence would lengthen the predicted rat glycosylase to 275 amino acids and extend the homologous region to the end of the human glycosylase (Fig. 3*C*).

The *AAG* Gene Maps to Human Chromosome 16. A panel of *Bam*HI-digested DNAs from 25 human–hamster somatic cell hybrids was probed with AAG cDNA. The human and hamster DNA hybridization patterns were easily distinguishable (Fig. 2*A*), and Table 1 shows that the *AAG* gene was only detected in hybrids 967 and 1079, whose common human chromosomes are 5 and 16. Since 20 of the other hybrids in the panel carried chromosome 5 it seemed likely that *AAG* maps to human chromosome 16. However, for one hybrid cell line, the results were unclear; hybrid 904 DNA did not hybridize to *AAG* even though 5% of the 904 cells are reported to contain human chromosome 16 (Table 1). Therefore, we confirmed the chromosome 16 assignment by probing DNA from a human–mouse hybrid cell line (GM10567) carrying chromosome 16 (and no other human chromosome) in 98% of the cells; the *AAG* gene was clearly present (Fig. 4). Because the single 6.0-kb mouse *Eco*RI band and one of the human *Eco*RI bands comigrated (Fig. 4, lanes 1 and 3), the identity of the human–mouse hybrid DNA loaded in lane 2 was confirmed by reprobing the same blot with a human chromosome 11-specific probe—namely, catalase (22); as expected for a probe not derived from human chromosome 16, the banding patterns in the human–mouse hybrid and the

A

```
                                                    CCTGGGCCCCCATGCCCGTGCAGCTCGC
-118  ACATATGTGGGGCAGAGCAGCCACCCTGCCCCCAGCAGCAGCCGTCCATCGTCAGACGT
 -59  GATCATTTCCTGAGGCCTCGAGTGTGTCAGGGTGTTTGTGCCTCATAACAACCCACAGG
  +1  ATG GTC ACC CCC GCT TTG CAG ATG AAG AAA CCA AAG CAG TTT TGC
       M   V   T   P   A   L   Q   M   K   K   P   K   Q   F   C
  46  CGA CGG ATG GGG CAA AAG AAG CAG CGA CCA GCT AGA GCA GGG CAG
       R   R   M   G   Q   K   K   Q   R   P   A   R   A   G   Q
  91  CCA CAC AGC TCG TCC GAC GCA GCC CAG GCA CCT GCA GAG CAG CCA
       P   H   S   S   S   D   A   A   Q   A   P   A   E   Q   P
 136  CAC AGC TCG TCC GAT GCA GCC CAG GCA CCT TGC CCC AGG GAG CGC
       H   S   S   S   D   A   A   Q   A   P   C   P   R   E   R
 181  TGC TTG GGA CCG CCC ACC ACT CCG GGC CCA TAC CGC AGC ATC TAT
       C   L   G   P   P   T   T   P   G   P   Y   R   S   I   Y
 226  TTC TCA AGC CCA AAG GGC CAC CTT ACC CGA CTG GGG TTG GAG TTC
       F   S   S   P   K   G   H   L   T   R   L   G   L   E   F
 271  TTC GAC CAG CCG GCA GTC CCC CTG GCC CGG GCA TTT CTG GGA CAG
       F   D   Q   P   A   V   P   L   A   R   A   F   L   G   Q
 316  GTC CTA GTC CGG CGA CTT CCT AAT GGC ACA GAA CTC CGA GGC CGC
       V   L   V   R   R   L   P   N   G   T   E   L   R   G   R
 361  ATC GTG GAG ACC GAG GCA TAC CTG TGG CCA GAG GAT GAA CCG GCC
       I   V   E   T   E   A   Y   L   G   P   E   D   E   P   A
 406  CAC TCA AGG GGT GGC CGG CAG ACC CCC CGC AAC CGA GGC ATG TTC
       H   S   R   G   G   R   Q   T   P   R   N   R   G   M   F
 451  ATG AAG CCG GGG ACC CTG TAC GTG TAC ATC ATT TAC GGC ATG TAC
       M   K   P   G   T   L   Y   V   Y   I   I   Y   G   M   Y
 496  TTC TGC ATG AAC ATC TCC AGC CAG GGG GAC GGG GCT TGC GTC TTG
       F   C   M   N   I   S   S   Q   G   D   G   A   C   V   L
 541  CTG CGA GCA CTG GAG CCC CTG GAA GGT CTG GAG ACC ATG CGT CAC
       L   R   A   L   E   P   L   E   G   L   E   T   M   R   H
 586  GTT CGC AGC ACC CTC CGG AAA GGC ACC GCC AGC CGT GTC CTC AAG
       V   R   S   T   L   R   K   G   T   A   S   R   V   L   K
 631  GAC CGC GAG CTC TGC AGT GGC CCC TCC AAG CTG TGC CAG GCC CTG
       D   R   E   L   C   S   G   P   S   K   L   C   Q   A   L
 676  GCC ATC AAC AAG AGC TTT GAC CAG AGG GAC CTG GCA CAG GAT GAA
       A   I   N   K   S   F   D   Q   R   D   L   A   Q   D   E
 721  GCT GTA TGG CTG GAG CGT GGT CCC CTG GAG CCC AGT GAG CCG GCT
       A   V   W   L   E   R   G   P   L   E   P   S   E   P   A
 766  GTA GTG GCA GCA GCC CGG GTG GGC GTC GGC CAT GCA GGG GAG TGG
       V   V   A   A   A   R   V   G   V   G   H   A   G   E   W
 811  GCC CGG AAA CCC CTC CGC TTC TAT GTC CGG GGC AGC CCC TGG GTC
       A   R   K   P   L   R   F   Y   V   R   G   S   P   W   V
 856  AGT GTG GTC GAC AGA GTG GCT GAG CAG GAC ACA CAG GCC TGA GCA
       S   V   V   D   R   V   A   E   Q   D   T   Q   A  ***
 901  AAGGGCCTGCCCAGACAAGATTTTTTAATTGTTTAAAAACCGAATAAATGTTTTATTTC
 960  TAGAAAAAAAAAAAAAAAAAA
```

B

```
H   1- MVTPALQMKK PKQFCRRMGQ KKQRPARAGQ PHSSSDAAQA PAEQPHSSSD
                          |  *  |**|| |||  |   |
R   1-           M PAPLSRKIGQ KKQQLAQSEQ QQTPKEKLSS

H  51- AAQAPCPRER CLGPPTTPGP YRSIYFSSPK GHLTRLGLEF FDQPAVPLAR
                            |||*  ||||||||   *  ||| |* |||||| |||
R  32-                    TPGL LRSIYFSSPE DRPARLGPEY FDQPAVTLAR

H 101- AFLGQVLVRR LPNGTELRGR IVETEAYLGP EDEPAHSRGG RQTPRNRGMF
       |||||||||| | *|||||||| |||||||||| ||| |||||| |||||||| |
R  66- AFLGQVLVRR LADGTELRGR IVETEAYLGP EDEAAHSRGG RQTPRNRGNF

H 151- MKPGTLYVYI IYGMYFCMNI SSQGDGACVL LRALEPLEGL ETMRHVRSTL
       |||||||||* |||||||*|* |||| ||||| |||||||||| ||||**|  |
R 116- MKPGTLYVYL IYGMYFCLNV SSQGAGACVL LRALEPLEGL ETMRQLRNSL

H 201- RKGTASRVLK DRELCSGPSK LCQALAINKS FDQRDLAQDE AVWLERGPLE
       ||     | ||  ||||| |||| ||||||   ||  |||||||||| ||      |
R 166- RKSWVGRSLK DRELCNGPSK LCQALARSKS FDQRDLAQDE AVCGWSMALW

H 251- PSEPAVVAAA RVGVGHAGEW ARKPLRFYVR GSPWVSVVDR VAEQDTQA
                  |      |
R 216- SPAASCGGGS RRYWPCRGMD TEALALLCPG QPMGQCGR
```

C

```
H 241- AVWLERGPLE PSEPAVVAAA RVGVGHAGEW ARKPLRFYVR GSPWVSVVDR
       |||||  ||||  |||||||||   | ||||||   |||||||  ||||||||||
R 206- AV*WLEHGPLE SSEPAVVAAA -AGIGHAGEW TQKPLRFYVQ GSPWVSVVDR

H 291- VAEQDTQA
       ||||  |
R 266- *VAEQMYQPQQ TACSDCSKVK
```

FIG. 3. Sequence analysis of the human 3-MeAde DNA glycosylase cDNA. (*A*) cDNA and predicted amino acid sequence. Three *Pst* I fragments of 0.3, 0.4, and 0.5 kb were subcloned from the pP5-3 cDNA insert into pTZ18R and pTZ19R and sequenced. The *Pst* I fragment order was confirmed by sequencing across the *Pst* I sites in pP5-3 with the appropriate oligonucleotide primers. Numbering is in relation to the first base of the ATG translation initiation codon at position +1. A putative polyadenylylation signal is underlined. (*B*) Comparison of the human and rat 3-MeAde DNA glycosylase amino acid sequences. The rat glycosylase sequence is taken from O'Connor and Laval (8). Exact matches are indicated by solid lines, similar amino acids by asterisks, and gaps by a dash. Note the sudden loss of identity near the carboxyl-terminal ends of the two glycosylases. (*C*) Slight modification of the rat glycosylase cDNA sequence extends homology. A −2 frameshift at nucleotide 745, and a +1 frameshift at nucleotide 777 in the rat 3-MeAde DNA glycosylase cDNA sequence (8) brings the rat glycosylase into exact register with the human glycosylase protein sequence.

mouse DNAs were similar to each other and different from human DNA.

None of the clinical disorders that have been mapped to chromosome 16 (23, 24) are suggestive of a DNA repair disorder; i.e., they are not known to involve spontaneous chromosome damage, a predisposition to cancer, or sensitivity to DNA-damaging agents such as chemotherapeutic drugs. However, another DNA repair gene—namely, *ERCC4*—has already been assigned to human chromosome 16 based upon the observation that this chromosome restores resistance to UV light to a UV-sensitive Chinese hamster ovary cell line belonging to complementation group 4 (25). It seems unlikely that the human 3-MeAde DNA glycosylase protects mammalian cells from UV-induced DNA damage, since it is unprecedented for 3-MeAde DNA glycosylases to provide UV resistance in either *E. coli* or *S. cerevisiae* cells (5). Therefore, it is improbable that *ERCC4* and *AAG* are allelic. However, it is formally possible that the substrates of this human 3-MeAde DNA glycosylase include UV-damaged bases. Until the *AAG* substrates have been identified, the *ERCC4* gene cloned and characterized, and the two genes more finely mapped on chromosome 16, the possibility that *AAG* and *ERCC4* encode the same DNA-repair enzyme can not be eliminated.

Until recently the isolation of human DNA repair genes by functional complementation was limited with respect to evolutionary distance; *E. coli* genes were cloned by suppressing *E. coli* repair defects, mammalian genes by suppressing mammalian repair defects, and so on. Clearly these boundaries can now be crossed, and eukaryotic DNA repair genes can be cloned by the suppression of *E. coli* DNA repair defects. So far this approach has been used to clone a yeast 3-MeAde DNA glycosylase gene (6), a human O^6-methylguanine DNA methyltransferase cDNA (26), a rat 3-MeAde DNA glycosylase cDNA (8), and now a human 3-MeAde DNA glycosylase cDNA (AAG). Moreover, this method of cloning mammalian DNA repair genes can be successful even when the cloned gene is not well expressed in *E. coli* (Fig. 1).

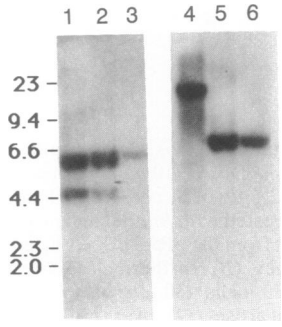

FIG. 4. Hybridization of the AAG cDNA to human chromosome 16. Eight micrograms of *Eco*RI-digested human DNA (lanes 1 and 4), DNA from the GM10567 human–mouse hybrid containing human chromosome 16 (lanes 2 and 5), and mouse DNA (lanes 3 and 6) were probed with a 0.5-kb *Pst* I fragment from the pP5-3 cDNA (lanes 1–3) and a catalase gene fragment (lanes 3–6).

Table 1. Distribution of the *AAG* gene and specific human chromosome in human–hamster hybrids

Hybrid	Human chromosome																							*AAG* gene
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y	
867	(+)	−	−	−	+	−	−	−	−	−	−	−	+	+	−	−	−	+	+	−	−	−	−	−	−
854	−	+	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
423	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
860	−	−	(+)	−	+	+	−	−	−	(+)	−	−	−	−	−	−	−	−	(+)	−	+	−	−	−	−
803	−	−	−	+	+	−	−	+	−	−	(+)	−	−	−	−	−	−	−	−	−	−	+	+	−	−
909	−	−	−	−	D	+	−	+	−	−	−	−	−	+	−	−	−	−	−	−	−	−	+	−	−
1006	−	−	−	(+)	+	−	+	+	−	−	−	−	+	−	+	−	−	−	+	−	+	−	−	−	−
811	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	+	+	−	−	−	−	−	−	−
967	−	−	−	−	+	−	−	+	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	+
734	−	−	−	−	+	−	−	−	+	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−
968	−	−	−	−	+	−	−	−	+	−	−	−	+	−	−	−	−	−	−	−	−	−	+	−	−
683	−	−	−	−	+	−	−	−	−	−	+	(+)	−	+	−	−	−	−	+	−	+	+	−	−	−
507	−	−	+	−	+	−	−	−	−	−	−	+	−	+	−	−	−	−	−	(+)	−	(+)	−	+	−
750	−	−	−	−	D	−	−	−	−	−	−	−	+	+	+	−	−	−	+	−	−	−	−	−	−
1099	+	−	−	−	D	−	−	−	−	−	−	−	+	−	−	−	−	−	+	−	+	+	−	−	−
324	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−
940	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−
983	−	−	−	−	+	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
937	+	−	−	−	+	−	−	−	−	−	−	−	−	+	+	−	+	−	−	−	+	−	−	−	−
1079	−	−	+	−	+	−	−	−	−	(+)	−	−	−	−	(+)	+	−	−	−	−	−	−	−	−	+
756	−	−	−	−	D	+	+	−	−	−	−	+	+	(+)	−	−	−	−	+	+	+	−	−	+	−
904	−	−	−	−	D	+	−	−	−	−	−	+	−	−	−	(+)	−	−	−	−	+	−	−	+	−
862	−	−	−	−	+	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
1049	−	−	−	−	+	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−
212	−	−	−	−	Dq	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−

Presence (+) or absence (−) of each human chromosome and the *AAG* gene in each hybrid line is indicated. A plus sign in parentheses indicates that <55% of the cells possess the given chromosome. The letter D indicates a deletion, and q indicates the long arm.

Cloning the AAG cDNA represents a critical step towards defining the *in vivo* role of the human 3-MeAde DNA glycosylase repair enzyme; the cloned gene can now be used to generate cell lines or animal strains that either overexpress or underexpress 3-MeAde DNA glycosylase. The AAG cDNA can be overexpressed from a strong promoter in human cell lines and transgenic mice. It can also be used to generate 3-MeAde DNA glycosylase-deficient human cell lines by antisense inhibition of enzyme production and targeted gene disruptions. Further, it may be possible to generate glycosylase-deficient mice by using the cloned mouse glycosylase gene and mouse embryonic stem cells. Such glycosylase mutants might reveal whether 3-MeAde repair prevents cell death, mutation, and chromosome damage induced by endogenous and exogenous alkylating agents, and whether this DNA repair function contributes to protecting animals against the carcinogenic effects of alkylating agents.

We thank John Cairns and Bruce Demple for their critical comments. This work was supported by American Cancer Society Research Grant NP448 and National Institute of Environmental Health Science Grant ES03926. L.S. and M.B. were supported respectively by an American Cancer Society Faculty Research Award and National Institutes of Health Training Grant CA 09078.

1. Friedberg, E. C. (1985) *DNA Repair* (Freeman, San Francisco).
2. Wallace, S. S. (1988) *Environ. Mol. Mutagen.* **12**, 431–477.
3. Lindahl, T., Sedgwick, B., Sekiguchi, M. & Nakabeppu, Y. (1988) *Annu. Rev. Biochem.* **57**, 133–158.
4. van Duin, M. & Hoeijmakers, J. H. J. (1989) *Ann. Ist. Super. Sanita* **25**, 131–142.
5. Thompson, L. H. (1988) in *DNA Repair, A Laboratory Manual of Research Procedures*, eds. Friedberg, E. C. & Hanawalt, P. C. (Dekker, New York), Vol. 3, pp. 115–132.
6. Chen, J., Derfler, B., Maskati, A. & Samson, L. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7961–7965.
7. Chen, J., Derfler, B. & Samson, L. (1990) *EMBO J.* **9**, 4569–4575.
8. O'Connor, T. & Laval, F. (1990) *EMBO J.* **9**, 3337–3342.
9. Volkert, M. R., Nguyen, D. C. & Beard, K. C. (1986) *Genetics* **112**, 11–26.
10. Prochownik, E. V., Markham, A. F. & Orkin, S. H. (1983) *J. Biol. Chem.* **258**, 8389–8394.
11. Talmadge, K., Stahl, S. & Gilbert, W. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3369–3373.
12. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
13. Rebeck, G. W., Coons, S., Carroll, P. & Samson, L. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3039–3043.
14. Samson, L. & Linn, S. (1987) *Carcinogenesis* **8**, 227–230.
15. Silhavy, T. J., Berman, M. L. & Enquist, L. W. (1984) *Experiments with Gene Fusions* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
16. Winston, F., Chumley, F. & Fink, G. R. (1983) *Methods Enzymol.* **101**, 211–227.
17. Singer, B. & Brent, T. P. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 856–860.
18. Gallagher, P. E. & Brent, T. P. (1982) *Biochemistry* **21**, 6404–6409.
19. Washington, W. J., Dunn, W. C., Generoso, W. M. & Mitra, S. (1988) *Mutat. Res.* **207**, 165–169.
20. Gerson, S. L., Miller, K. & Berger, N. A. (1985) *J. Clin. Invest.* **76**, 2106–2114.
21. Gerson, S. L., Trey, J. E., Miller, K. & Benjamin, E. (1987) *Cancer Res.* **47**, 89–95.
22. Bruns, G. A. P., Glaser, T., Gusella, J. F., Housman, D. & Orkin, S. H. (1984) *Am. J. Hum. Genet.* **36**, 25S (abstr.).
23. Reeders, S. T. & Hildebrand, C. E. (1989) *Cytogenet. Cell Genet.* **51**, 299–318.
24. Harper, P. S., Frezal, J., Ferguson-Smith, M. A. & Schinzel, A. (1989) *Cytogenet. Cell Genet.* **51**, 563–611.
25. Siciliano, M. J., Bachinsky, L., Dolf, G. Carrano, A. V. & Thomson, L. H. (1987) *Cytogenet. Cell Genet.* **46**, 691–692.
26. Tano, K., Shiota, S., Collier, J., Foote, R. S. & Mitra, S. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 686–690.