JOURNAL OF BACTERIOLOGY, July 1996, p. 3885–3892 Vol. 178, No. 13
0021-9193/96/$04.00+0

Cloning and Sequencing a Human Homolog (*hMYH*) of the *Escherichia coli mutY* Gene Whose Function Is Required for the Repair of Oxidative DNA Damage

MALGORZATA M. SLUPSKA,[1] CLAUDIA BAIKALOV,[1] WENDY M. LUTHER,[1] JU-HUEI CHIANG,[1] YING-FEI WEI,[2] AND JEFFREY H. MILLER[1]*

Department of Microbiology and Molecular Genetics and the Molecular Biology Institute, University of California, Los Angeles, California 90024,[1] and Human Genome Sciences, Rockville, Maryland 20850[2]

Received 22 January 1996/Accepted 29 April 1996

We have cloned the human *mutY* gene (*hMYH*) from both genomic and cDNA libraries. The human gene contains 15 introns and is 7.1 kb long. The 16 exons encode a protein of 535 amino acids that displays 41% identity to the *Escherichia coli* protein, which provides an important function in the repair of oxidative damage to DNA and helps to prevent mutations from oxidative lesions. The human *mutY* gene maps on the short arm of chromosome 1, between p32.1 and p34.3.

The study of bacterial and yeast repair genes and their involvement in preventing mutations has facilitated the understanding of mutagenesis and repair in humans. A stunning example of this is the recent demonstration that inherited susceptibility to hereditary nonpolyposis colon cancer (HNPCC) is due to the inheritance of a defective copy of one of the human homologs of the bacterial mismatch repair system (4, 6, 8, 19). Defects in either the *Escherichia coli mutH*, *-L*, or *-S* or the *uvrD* gene lead to a nonfunctional mismatch repair system and an increased rate of spontaneous mutations, the "mutator" phenotype (16). In humans, a mutator effect, easily detectable by observing instability of microsatellite repeats (8, 9), is also seen to result from defects in the mismatch repair system. This has led to renewed interest in characterizing mutator bacteria that might define additional repair systems and in finding their counterparts in humans.

We have described two mutator genes in *E. coli*, the *mutY* and the *mutM* genes (5, 17), which work together to prevent mutations from certain types of oxidative damage, dealing in particular with the oxidized guanine lesion 8-oxodG (13). Figure 1 (see also reference 14) summarizes the concerted action of the enzymes encoded by these two genes, both of which are glycosylases. The MutM protein removes 8-oxodG from the DNA, and the resulting a purinic site is repaired to restore the GC base pair. Some lesions are not repaired before replication, which results in a GC-to-TA transversion at the next round of replication because polymerases involved in DNA replication incorporate A across from 8-oxoG between 5-fold and 200-fold more frequently than they incorporate C (22). However, the MutY protein removes the A across from 8-oxodG and repair synthesis restores a C most of the time (since polymerases involved in repair insert C in preference to A across from 8-oxoG [22]), allowing the MutM protein another opportunity to repair the lesion. In accordance with this, mutators lacking either the MutM or the MutY protein have an increase specifically in the GC-to-TA transversion (5, 17), and cells lacking both enzymes have an enormous increase in this base substitution (13). A third protein, the product of the *mutT* gene, prevents the incorporation of 8-oxodGTP by hydrolyzing the oxidized triphosphate back to the monophosphate (11), preventing AT-to-CG transversions.

The human homolog of the *E. coli mutT* gene has been cloned and sequenced (20). Here we describe the characterization of the human homolog of the *mutY* gene (*hMYH*). The gene is 7.1 kb long and contains 15 introns, most of which are less than 200 bp in length. The exons, revealed by comparing the genomic sequence with that from HeLa, brain, and testis cDNA clones, encode a 535-amino-acid protein which has 41% identity to the shorter *E. coli* protein. Using in situ hybridization, we have mapped the human *mutY* gene to the short arm of chromosome 1, between p32.1 and p34.3.

MATERIALS AND METHODS

cDNA and genomic libraries. cDNA libraries used in cloning of human *mutY* were as follows: brain (Human Genome Sciences); brain, cerebellum; Lambda ZAP library (catalog no. 935 201; Stratagene, La Jolla, Calif.), HeLa cell line S3, grown to semiconfluency; and Uni ZAP XR library (catalog no. 937216; Stratagene). We also used a genomic pBAC library from human fibroblasts, kindly supplied by M. Simon. All media and genetic manipulation were as described elsewhere (15).

DNA synthesis and sequencing. Oligonucleotides were synthesized on a Beckman oligo1000 DNA synthesizer by using solid-phase cyanoethyl phosphoramidite chemistry. All oligonucleotides were deprotected in ammonium hydroxide and used without further purification.

DNA sequencing was done by using [α-^{32}P]dATP and a SequiTherm Cycle Sequencing Kit (Epicentre Technologies, Madison, Wis.) with reagents supplied by the manufacturer. For sequencing cDNA clones, we employed PCR conditions described by the manufacturer; for sequencing genomic clones (~100 kb), we performed one cycle at 94°C for 5 min followed by 25 cycles of 94°C for 45 s, 52°C for 30 s, and 72°C for 5 min.

PCR amplification of cDNA clones and 5′ rapid amplification of cDNA ends (RACE) reactions. Standard PCR to amplify cDNA clones was performed with 50-μl reaction mixtures containing 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM each deoxynucleoside triphosphate, 0.4 mM each primer, 1 U of *Taq* DNA polymerase (GibcoBRL, Life Technologies Inc., Gaithersburg, Md.), and 1 μl of cDNA library (~10^{10} PFU/ml). Two different amplification cycles on a PTC-100 thermal cycler (MJ Research, Inc., Watertown, Mass.) were used. For PCR of cDNA clones, we used 30 cycles of 94°C for 30 s, 52°C for 30 s, and 72°C for 1 min, followed by 10 min at 72°C. To determine the size of bigger introns, we used 25 cycles of 94°C for 45 s, 52°C for 30 s, and 72°C for 5 min. For all reactions, we used hot start: 5 min at 94°C followed by a decrease to 80°C when *Taq* polymerase was added. A 5-μl amount of PCR mixture was analyzed by agarose gel electrophoresis followed by ethidium bromide staining.

To find cDNA clones of human *mutY*, we performed two-step PCR. In the first step we used primer 23 on the human *mutY* sequence (see Table 1 and Fig. 4) and primer L1 on the Lambda vector (Table 1). The resulting PCR product was purified by using the QIAquick PCR Purification Kit (Qiagen Inc., Chatsworth, Calif.) and eluted with 50 μl of 10 mM Tris-HCl buffer (pH 8.0), and 2 μl of this preparation was used as the DNA template in the second PCR with primers IcD

* Corresponding author.

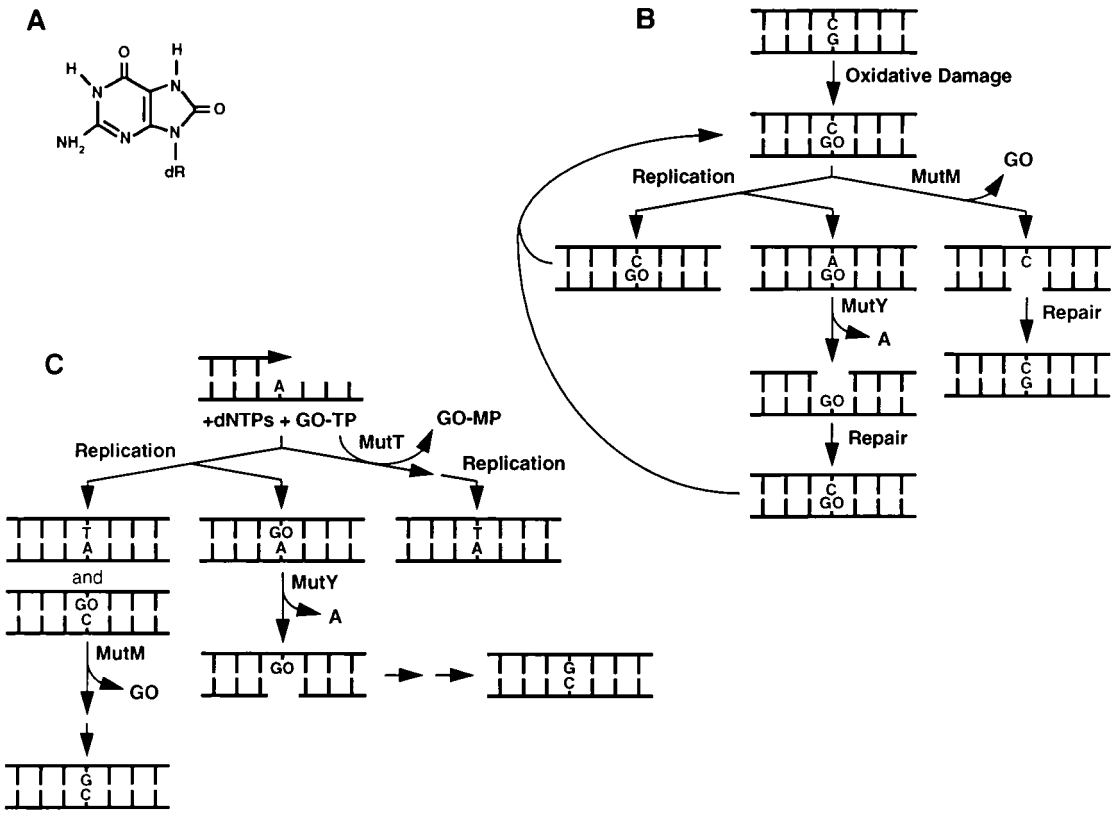


FIG. 1. The GO system. (A) 7,8-Dihydro-8-oxoguanine (8-hydroxyguanine). This is the structure of the predominant tautomeric form of the GO lesion. (B) Oxidative damage can lead to GO lesions in DNA. The GO lesions can be removed by MutM protein, and subsequent repair can restore the original GC base pair. If the GO lesion is not removed before replication, translesion synthesis by replicative DNA polymerases is frequently inaccurate, leading to the misincorporation of A opposite the GO lesion. MutY removes the misincorporated adenine from the A-GO mispairs. Repair polymerases are much less error prone during translesion synthesis and can lead to a C-GO pair—a substrate for MutM. (C) Oxidative damage can also lead to 8-oxo-dGTP. MutT is active on 8-oxodGTP and hydrolyzes it to 8-oxodGMP, effectively removing the triphosphate from the deoxynucleotide pool; otherwise, inaccurate replication could result in the misincorporation of 8-oxodGTP opposite template A residues, leading to A-GO mispairs. MutY could be involved in the mutation process because it is active on the A-GO substrate and would remove template A, leading to the AT→CG transversions that are characteristic of a *mutT* strain. The 8-oxodGTP could also be incorporated opposite template cytosines, resulting in a damaged C-GO pair that could be corrected by MutM. dNTP, deoxynucleoside triphosphates.

and 58 or primers I7 and 23 (Table 1). Resulting cDNA bands were agarose purified by using Gelase (Epicentre Technologies) and the QIAEX II DNA purification kit (Qiagen Inc.) and then cloned into pCR-Script SK(+) vector by using the pCR-Script SK(+) Cloning Kit (Stratagene). Cloned fragments were sequenced.

To find the 5′ end of the cDNA of the human *mutY* gene, we employed a 5′ RACE kit (GibcoBRL, Life Technologies Inc.) with poly(A)$^+$ RNA from human testis tissue (Clontech Laboratories, Inc., Palo Alto, Calif.), using reagents and protocols supplied by the manufacturer. As nested primer no. 1, we used primer 23 (Table 1); for PCR amplification, we used primers 58 and 64 (Table 1). The resultant DNA bands were agarose purified and cloned into pCR-Script SK(+) as described above.

Cloning the genomic clone of human *mutY* for fluorescence in situ hybridization (FISH) mapping. A ~40-μg amount of the pBAC clone (~100 kb) containing the human *mutY* gene was digested with 0.5 U of *Sau*3AI for 5 min at 37°C by using enzyme and reaction buffer obtained from New England BioLabs (Beverly, Mass.). Digested DNA was purified with the QIAEX II DNA purification kit (Qiagen Inc.) and ligated to SuperCos1 cosmid vector (Stratagene) digested and dephosphorylated according to the protocols supplied by the manufacturer. Ligated DNA was then packed by using Gigapack III Gold Packing Extract (Stratagene), and the resultant cosmids were screened by hybridization with the PCR-amplified segments of *mutY* cDNA.

Hybridization. Human *mutY* cDNA fragments for use as hybridization probes were obtained by PCR amplification with primers 22 and 23 (for screening the genomic library) and primers 22 and 23 or IcD and 58 (for screening cosmid libraries). All PCR-amplified DNA fragments were agarose purified as described above and α-^{32}P labeled by using the Prime-It II kit from Stratagene. In screening genomic libraries for background radiation, we also labeled pFOS DNA with ^{35}S using the same Prime-It II kit. Prehybridization solution contained 1 M NaCl, 50 mM Tris-HCl (pH 8.0), 5 mM EDTA (pH 8.0), 1% sodium dodecyl sulfate (SDS), and 10% dextran sulfate. The prehybridization solution was preheated to 65°C, and then filters were added and the mixture was agitated at 65°C for 2 h at 100 rpm. At that point, the DNA probes were boiled for 10 min together with blocking human placental DNA (10 mg/ml) (Sigma, St. Louis, Mo.) and annealed at 65°C for 30 min. pFOS probe was boiled for 5 min, and then all probes were added to prehybridized filters. Hybridization was carried out for 18 to 20 h at 65°C. After hybridization, filters were washed once in 1× SSC (0.15 M NaCl plus

TABLE 1. Primers used for amplification of *hMYH* cDNA fragments

Primer[a]	Sequence
Exons 1–10	
64-N	5′TCCTCTGAAGCTTGAGGAGCCTCTAGAACT3′
58-C	5′TAGCTCCATGGCTGCTTGGTTGAAA3′
Exons 2–10	
IcD-N	5′GCCATCATGAGGAAGCCACGAGCAG3′
58-C	5′TAGCTCCATGGCTGCTTGGTTGAAA3′
Exons 6–16	
I7-N	5′TTGACCCGAAACTGCTGAATAG3′
23-C	5′CAGTGGAGATGTGAGACCGAAAGAA3′
Exons 10–16	
22-N	5′CAGCCCGGCCAGGAGATTTCAACCA3′
23-C	5′CAGTGGAGATGTGAGACCGAAAGAA3′
L1 on Lambda ZAP	5′CCCTCACTAAAGGGAACAAAAGCTGG3′

[a] N indicates the primer on the 5′ side of the gene; C indicates the primer on the 3′ side.

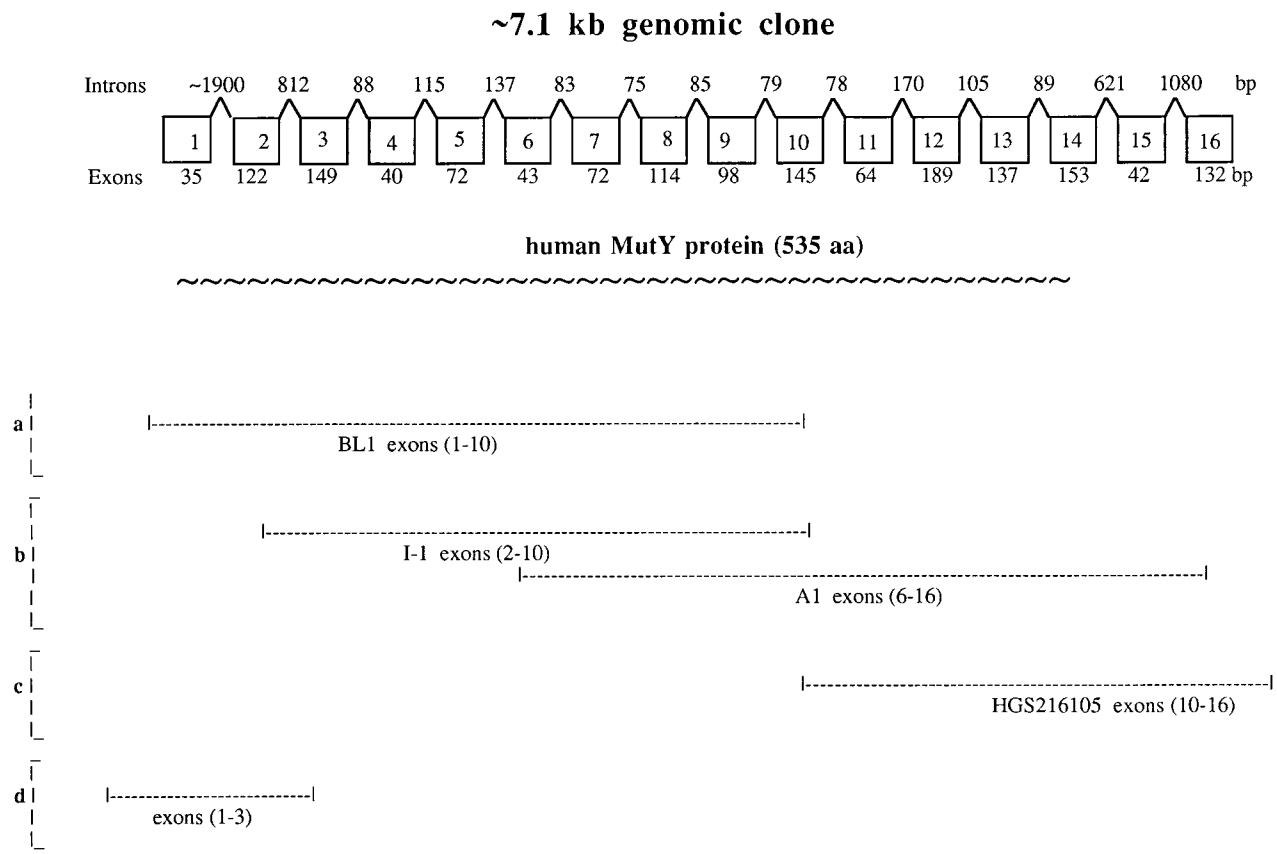


FIG. 2. Organization of the human *mutY* locus and cDNA clones. The boxes containing numbers 1 to 16 represent individual *hMYH* exons. The size of each exon is given below each box, and the size of each intron is given above the region between individual pairs of exons. The lines below the gene represent each of the cDNA clones obtained. cDNA clones are labeled with an identification name and the identification number of each exon contained in the clone. Lines: a, human testis *MYH* cDNA obtained with 5′ RACE kit; b, HeLa cell library (Stratagene); c, brain library (Human Genome Sciences); d, GenBank.

0.015 M sodium citrate)–0.17% SDS–0.05% sodium PP_i at room temperature for 15 min and twice in 0.1× SSC–0.1% SDS at 65°C for 15 min.

FISH mapping. DNA of the cosmid clone used for FISH mapping was purified by using the QIAgen Plasmid Purification Kit (Qiagen Inc.), and 1 μg was labeled by nick translation in the presence of biotin-dATP by using the BioNick Labeling Kit (GibcoBRL, Life Technologies Inc.). Biotinylation was detected with the GENE-TECT Detection System (Clontech Laboratories, Inc.). In situ hybridization was performed on slides by using the ONCOR Light Hybridization Kit (ONCOR, Gaithersburg, Md.) to detect single-copy sequences on metaphase chromosomes. Peripheral blood of healthy donors was cultured for 3 days in RPMI 1640 supplemented with 20% fetal calf serum, 3% phytohemagglubinin, penicillin, and streptomycin; synchronized with 10^{-7} M methotrexate for 17 h; and washed twice with unsupplemented RPMI. Cells were incubated with 10^{-3} M thymidine for 7 h. Growth of the cells was arrested in metaphase after 20 min of incubation with colcemid (0.5 μg/ml), and then cells were subjected to hypotonic lysis in 75 mM KCl for 15 min at 37°C. Cell pellets were then spun out and fixed in Carnoy's fixative (methanol-acetic acid, 3:1). Metaphase spreads were prepared by adding a drop of the suspension onto slides and air dried. Hybridization was performed by adding 100 ng of probe suspended in 10 ml of hybridization mix (50% formamide, 2× SSC, 1% dextran sulfate) with blocking human placental DNA (1 μg/ml). The probe mixture was denatured for 10 min in a 70°C water bath and incubated for 1 h at 37°C, before being placed on a prewarmed (37°C) slide, which was previously denatured in 70% formamide–2× SSC at 70°C, and dehydrated in an ethanol series, chilled to 4°C. Slides were incubated for 16 h at 37°C in a humidified chamber. Slides were washed in 50% formamide–2× SSC for 10 min at 41°C and in 2× SSC for 7 min at 37°C. Hybridization probe was detected by incubation of the slides with fluoroisothiocyanate-avidin (ONCOR) according to the manufacturer's protocol. Chromosomes were counterstained with propidium iodine suspended in mounting medium. Slides were visualized by using a Leitz ORTHOPLAN 2-epifluorescence microscope, and five computer images were taken, by using an Imagenetics computer and a MacIntosh printer.

RESULTS

Cloning the human *mutY* gene. We have screened the Human Genome Sciences human cDNA sequence database for fragments encoding protein sequences homologous to the *E. coli* MutY protein. We detected one clone from a human brain cDNA library with this property. The initial expressed sequence tag (EST) clone used in the study was discovered by scientists at the Institute for Genomic Research using established EST methods (1, 2). This clone is part of a larger EST project (3). We extended the sequence of the cDNA fragment, which was 1,200 bp long and contained the 3′ end of the *mutY* coding sequence.

Using primers 22 and 23 (Table 1), we PCR amplified the 800-bp fragment of cDNA used as a probe for hybridization with a human pBAC library, with an average insert size of 100 to 300 kb. We detected four clones which hybridized to the probe: 79B8 (2CH), 79D11 (3CH), 124B11 (4CH), and 815G2 (7CH). For further analysis, we used clone 79B8 (2CH) and we sequenced a segment of it using primer extension. By examining protein homologies and applying the rules for intron-exon boundaries (18, 21), we were able to predict the positions of most of the introns and exons.

By designating primers from different predicted exons, we screened two cDNA libraries from human brain tissue (Stratagene and Human Genome Sciences) and one from HeLa cells (Stratagene). We found that only the HeLa cell library con-

FIG. 3. Nucleotide sequence of human *mutY* and its deduced amino acid sequence. The nucleotide sequence of *hMYH* is shown, with numbering relative to the A of the ATG translational start site (+1). Amino acid sequence is shown below in single-letter code and is also numbered in the margin. Bold letters indicate the differences found in different libraries (see Results).

Non coding region (Exon Coordinates) Non coding region

Exon 1 (1 to 35)
aaactgcgccatcgtcactgtcggcggccATGACACCGCTCGTCTCCCGCCTGAGTCGTCTGTGGgtacgcgtggacttgcggtccgtctcctg

Exon 2 (36 to 157)
tttggctgggtctttttgtttcagGGCCATCATGAGGAAGCCAC.....CAAGCCTTCTGCCTGTGATGgtaaggaactaggttgtggcccaa

Exon 3 (158 to 306)
gtgtcctggggccccagcagGCCTGGCCAGGCAGCCGGAA.....ACCTACCATGGAGAAGACGGgtaggcaggcgaggagcagg

Exon 4 (307 to 346)
tcatctggggttgcattgacagGCAGAAGATGAGATGGACCTGGACAGGCGGGCATATGCTGgtcagtacatctccttgagagcagggcc

Exon 5 (347 to 420)
catgccaaccccttttccccagTGTGGGTCTCAGAGGTCATG.....TATACCGGATGGATGCAGgtgactccaggggaggaagggaagg

Exon 6 (421 to 462)
tttctgcctgcctgtggctatagAAGTGGCCTACACTGCAGGACCTGGCCAGTGCTTCCCTGGAGgtgagagccaccctagggtagggg

Exon 7 (463 to 534)
tgacctctgatcctacccacagGAGGTGAATCAACTCTGGGC.....TGCAGGAGGGAGCTCGGAAGgtaaggggatggcaggagggt

Exon 8 (535 to 648)
ttataatcttgcactccaatcagGTGGTAGAGGAGCTAGGGGG.....CCTCTATCGCCTTTGGCCAGgtgatctcacagcccacccccactt

Exon 9 (649 to 746)
gctaactctttggcccctctgtgccagGCAACCGGTGTGGTGGATGG.....CTTGTTTCCCAGCAGCTCTGgtaggatgttggggtaacaagggt

Exon 10 (747 to 891)
gcttcacagcagtgttccctctttagGGGTCTAGCCCAGCAGCTGG.....TGTGCCGGGCACGCCAGAGAgtaagcctactggggaagggg

Exon 11 (892 to 955)
cactcaaccctgtgcctctcagGTGGAGCAGGAACAGCTCTT.....TCCTGACGTGGAGGAGTGTGgtgagcaccaaacctagccccca

Exon 12 (956 to 1144)
gttcggggtactccgttcccagCTCCCAACACTGGACACTGC.....GGTGCAGAGGCCCAACTCAGgtacctggatactgggcgtggag

Exon 13 (1145 to 1281)
gcctggctgccctccctctcagGTCTGCTGGCAGGACTGTGG.....ACCTCCGGCACCTTGGGGAGgtaagtgagcagcggaatagcc

Exon 14 (1282 to 1434)
aacccctgaccttcctccagGTTGTCCACACCTTCTCTCA.....GTTTCCACCGCCATGAAAAGgcactacctttgttgtctttgttgtacttc

Exon 15 (1435 to 1476)
tcacctccctgtcttcttgtctagGTTTTCCGTGTGTATCAGGGCCAACAGCCAGGGACCTGTATGgtaagtctcctaggcctctcccaaccg

Exon 16 (1477 to 1605)
ctccctccttccatttttcacagGGTTCCAAAAGGTCCCAGGT.....TCAACAGTGCAGCCCAGTGAcacctctgaaagcccccattcccctga

FIG. 4. Sequences of the exon-intron boundaries. Sequences of introns are given in lowercase, and those of exons are in uppercase. The numbers in parentheses above the exon sequences are nucleotide coordinates of the exon sequences, assuming that the A of ATG is nucleotide 1. Additional intron sequence (sequences of whole introns except for intron 1) is available on request.

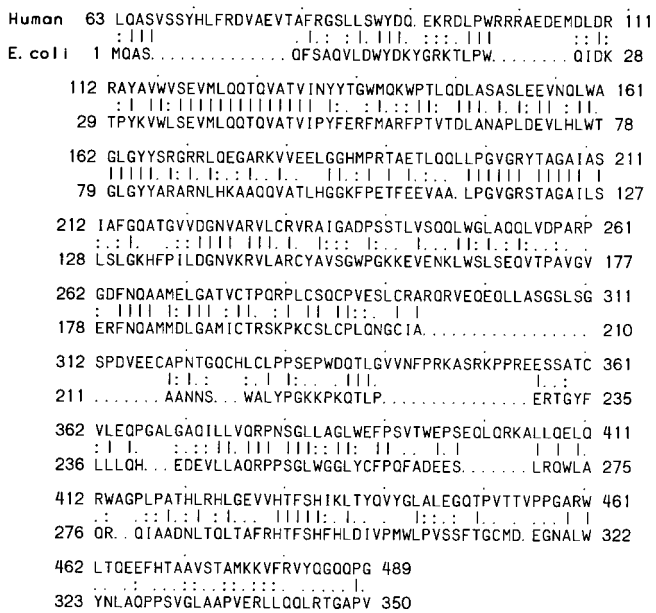


FIG. 5. Comparison of predicted amino acid sequences of human and *E. coli* MutY proteins. A "|" indicates identical amino acids, "." and ":" indicate similar amino acids, and a "..." represents a gap that has been introduced to optimize the homology. The comparison was done by using Bestfit (7).

tained a detectable level of *mutY* cDNA. Using primers IcD and 58 and primers I7 and 23 (Table 1), we amplified two DNA segments, which were then sequenced. The 5′ terminus was found by employing the 5′ RACE kit (GibcoBRL) with poly(A)$^+$ RNA from human testis tissue (Clontech) with primers 64 and 58. Sequencing overlapping clones allowed the verification of the gene structure, which is shown in Fig. 2. Additional confirmation of the sequence of the extreme 5′ end of the human *mutY* cDNA came from the sequence of the EST07222 clone retrieved from GenBank because of the overlap with the genomic sequence of the human *mutY* gene. Subsequently, we retrieved two additional sequence stretches of human *mutY* cDNA from GenBank: EST56698 at the 5′ end and yj560.s1 at the 3′ end.

The human *mutY* gene contains 15 introns ranging in size from 42 to approximately 1,500 bp, although most of the introns are less than 200 bp. The entire gene is 7,100 bp and encodes a protein, MutY, of 535 amino acids. Figure 2 shows the clones used in determining the gene structure, and the sequence of the *mutY* cDNA is given in Fig. 3. We found three differences between DNAs from different libraries. Two of them are silent mutations: C at position 195 in the genomic sequence is changed to T in the cDNA from testis tissue, and C at position 558 in the genomic sequence is changed to T in the cDNA from testis tissue. One difference changes the amino acid histidine to glutamine (C at position 972 in the genomic DNA is changed to G in the cDNA from brain tissue). All the changes are indicated in Fig. 3. Figure 4 depicts the intron-exon boundaries.

The human MutY protein shows 41% identity to the *E. coli* MutY protein, as determined by the Bestfit program (7). These two protein sequences are shown in Fig. 5. Note the extensive homology near the beginning of the *E. coli* protein, which is characterized by a run of 14 identical amino acids in a stretch in which 16 of 17 amino acids are identical. Table 2 reports the comparison of the MutY sequence with other known MutY proteins.

FISH mapping. We used a genomic clone containing the first 15 exons of the human *mutY* gene (Fig. 2) for mapping the gene on the human chromosome by FISH. Clone 156 was obtained by *Sau3*AI digestion of the 79B8 (2CH) clone from the genomic library followed by ligation to the SuperCos1 cosmid vector (Stratagene) and packaging of ligated DNA by using Gigapack III Gold Extract (Stratagene). The cosmid clones from subcloning of 2CH were screened by hybridization to the PCR-amplified segments of human *mutY* cDNA. Cosmid 156 hybridized to two segments of *mutY* cDNA obtained with primers 22 and 23 and primers IcD and 58. It contained the 20-kb insert with 15 exons and 14 introns from the 5′ end of the *mutY* gene. By using clone 156, the human *mutY* gene was mapped on the short arm of chromosome 1 (Fig. 6).

DISCUSSION

The GO system in bacteria prevents mutations arising from the oxidation product 8-oxodG (13, 14). Of the three components of this system, the MutY, MutM, and MutT proteins, two have now been characterized in humans. The *mutT* gene was previously cloned, and the protein was expressed in bacteria (20). Here we report the cloning and sequencing of the human *mutY* gene and show that it maps on the short arm of chromosome 1, at p32.1 to p34.3. The human *mutY* gene translated from our clone encodes 535 amino acids, which is in good agreement with the size of a polypeptide detected in HeLa cell material that cross-reacted with antibody against *E. coli* MutY protein (12).

Because cells lacking the human mismatch repair system are mutators and this defect leads to an increased susceptibility to colon cancer (4, 6, 8, 19), it is possible that any repair defect that generates a mutator phenotype might also lead to cancer susceptibility. Therefore, it becomes crucially important to characterize new repair systems and to examine the consequences of defects in these systems. The system repairing 8-oxodG in bacteria clearly has its counterpart in humans. A detailed comparison of each enzyme activity awaits high-level expression and purification of both the MutY and, subsequently, the MutM proteins. We are using the human *mutY* clone to construct a knockout mouse lacking the *mutY* gene. Will heterogenotes with one good copy have an increased susceptibility to cancer? Ultimately, it will be interesting to compare the properties of a double-knockout mouse, defective in both copies of the *mutY* and *mutM* genes, with those of *mutY mutM* bacteria, which have an enormous mutation rate (13).

The human *mutY* gene sequence can also be used to screen selected tumor lines for defects in this gene. Some tumor lines have intact mismatch repair systems yet are still mutators (16a). In fact, it has been argued that a mutator phenotype may be required for multistage carcinogenesis (9, 10).

ACKNOWLEDGMENTS

We thank workers at The Institute for Genomic Research sequencing facility for the initial sequencing of EST clones and workers at the

TABLE 2. Homology of hMYH protein with sequences of known MutY proteins[a]

Species	Quality of homology	% Similarity	% Identity
Escherichia coli	234.2	61.032	40.974
Salmonella typhimurium	245.0	59.249	38.338
Haemophilus influenzae	235.2	55.467	36.5533

[a] Homology data were obtained by Bestfit (7).

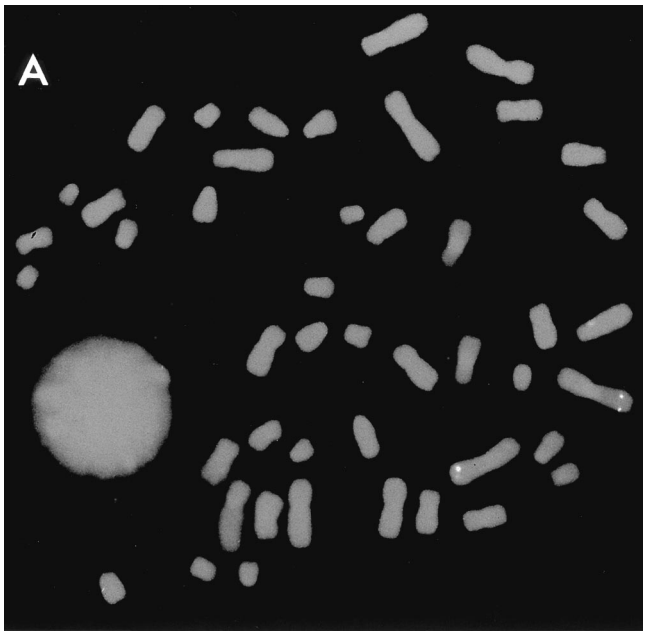
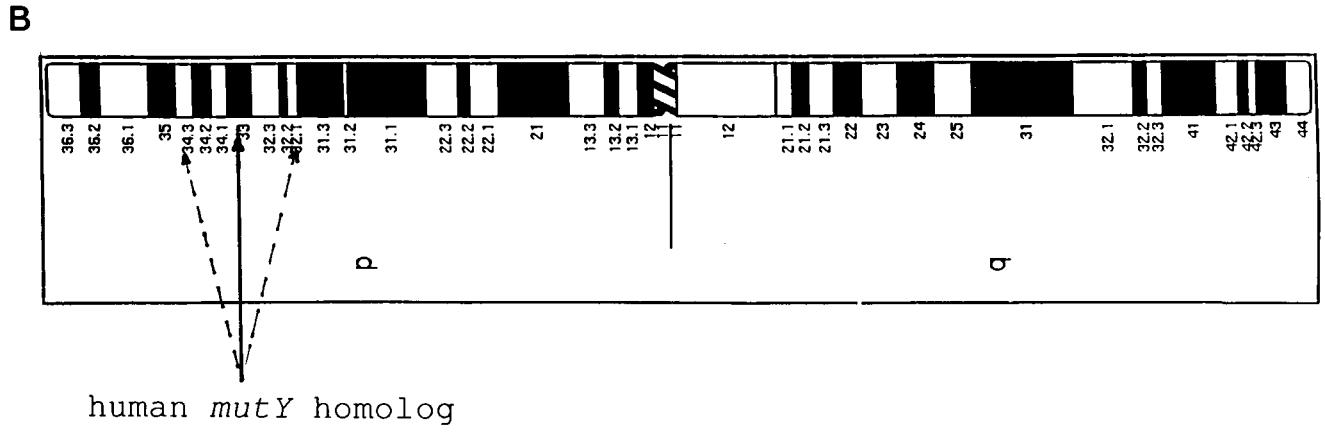


FIG. 6. Chromosome localization of the *hMYH* determined by FISH. (A) Example of hybridization signals of twin dots on chromosome 1; (B) ideogram of human chromosome 1 showing localization of the *hMYH* gene on 1p32 to 1p34 (arrows) based on analysis of five metaphases with twin dots.



Human Genome Sciences, Inc., sequencing core facility, where sequencing of related constructs was carried out. We thank Ivana Klisak, in collaboration with Richard Gatti at UCLA Medical School, for carrying out the FISH mapping. We are grateful to Richard Kolodner for helpful discussions and technical advice.

This work was supported by a grant to J.H.M. from the National Institutes of Health (GM 32184).

REFERENCES

1. **Adams, M. D., M. Dubnick, A. R. Kerlavage, R. Moreno, J. M. Kelley, T. R. Utterback, J. W. Nagle, C. Fields, and J. C. Venter.** 1992. Sequence identification of 2,375 human brain genes. Nature (London) **355:**632–634.
2. **Adams, M. D., J. M. Kelley, J. D. Gocayne, M. Dubnick, M. H. Polymeropoulos, H. Xiao, C. R. Merril, A. Wu, B. Olde, R. F. Moreno, R. Kerlavage, W. R. McCombie, and J. C. Venter.** 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. Science **252:**1651–1656.
3. **Adams, M. D., A. R. Kerlavage, R. D. Fleischmann, R. A. Fuldner, C. J. Bult, N. H. Lee, E. F. Kirkness, K. G. Weinstock, J. D. Gocayne, O. White, et al.** 1995. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature (London) **377**(Suppl. 6547)**:**3–174.
4. **Bronner, C. E., S. M. Baker, P. T. Morrison, G. Warren, L. G. Smith, M. K. Lescoe, M. Kane, C. Earabino, J. Lipford, A. Lindblom, P. Tannergard, R. J. Bollag, A. R. Godwin, D. C. Ward, M. Nordenskjold, R. Fishel, R. Kolodner, and R. M. Liskay.** 1994. Mutation in the DNA mismatch repair gene homologue *hMLH1* is associated with hereditary nonpolyposis colon cancer. Nature (London) **368:**258–261.
5. **Cabrera, M., Y. Nghiem, and J. H. Miller.** 1988. *mutM*, a second mutator locus in *Escherichia coli* that generates G · C → T · A transversions. J. Bacteriol. **170:**5405–5407.
6. **Fishel, R., M. K. Lescoe, M. R. S. Rao, N. G. Copeland, N. A. Jenkins, J. Garber, M. Kane, and R. Kolodner.** 1993. The human mutator gene homolog *MSH2* and its association with hereditary nonpolyposis colon cancer. Cell **75:**1027–1038.
7. **Genetics Computer Group.** 1991. Program manual for the GCG package, version 7. Genetics Computer Group, Madison, Wis.
8. **Leach, F. S., N. C. Nicolaides, N. Papadopoulos, B. Liu, J. Jen, R. Parsons, P. Petlomäki, P. Sistonene, L. A. Aaltonen, M. Nyström-Lahti, X.-Y. Guan, J. Zhang, P. S. Meltzer, J.-W. Yu, F.-T. Kao, D. J. Chen, K. M. Cerosaletti, R. E. Fournier, S. Rodd, T. Lewis, R. J. Leach, S. L. Naylor, J. Weissenbach, J.-P. Mecklin, H. Järvinen, G. M. Petersen, S. R. Hamilton, J. Green, J. Jass, P. Watson, H. T. Lynch, J. M. Trent, A. de la Chapelle, K. W. Kinzler, and B. Vogelstein.** 1993. Mutations of a *mutS* homolog in hereditary nonpolyposis colorectal cancer. Cell **75:**1215–1225.
9. **Loeb, L. A.** 1994. Microsatellite instability: marker of a mutator phenotype in cancer. Cancer Res. **54:**5059–5063.
10. **Loeb, L. A.** 1991. A mutator phenotype may be required for multistage carcinogenesis. Cancer Res. **51:**3075–3079.
11. **Maki, H., and M. Sekiguchi.** 1992. MutT protein specifically hydrolyzes a potent mutagenic substrate for DNA synthesis. Nature (London) **355:**273–275.
12. **McGoldrick, J. P., Y. C. Yeh, M. Solomon, J. M. Essigmann, and A. L. Lu.**

1995. Characterization of a mammalian homolog of the *Escherichia coli* MutY mismatch repair protein. Mol. Cell. Biol. **15:**989–996.
13. **Michaels, M. L., C. Cruz, A. P. Grollman, and J. H. Miller.** 1992. Evidence that *mutM* and *mutY* combine to prevent mutations by an oxidatively damaged form of guanine in DNA. Proc. Natl. Acad. Sci. USA **89:**7022–7025.
14. **Michaels, M. L., and J. H. Miller.** 1992. The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). J. Bacteriol. **174:**6321–6325.
15. **Miller, J. H.** 1992. A short course in bacterial genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Modrich, P.** 1991. Mechanisms and biological effects of mismatch repair. Annu. Rev. Genet. **25:**229–253.
16a. **Modrich, P., and R. Kolodner.** Personal communication.
17. **Nghiem, Y., M. Cabrera, C. G. Cupples, and J. H. Miller.** 1988. The mutY gene: a mutator locus in *Escherichia coli* that generates G:C to T:A transversions. Proc. Natl. Acad. Sci. USA **85:**9163–9166.
18. **Padgett, R. A., P. J. Grabowski, M. M. Konarska, S. Seiler, and P. A. Sharp.** 1986. Splicing of messenger RNA precursors. Annu. Rev. Biochem. **55:**1119–1150.
19. **Papadopoulos, N., N. C. Nicolaides, Y. F. Wei, S. M. Ruben, K. C. Carter, C. A. Rosen, W. A. Haseltine, R. D. Fleischmann, C. M. Fraser, and M. D. Adams.** 1994. Mutation of a *mutL* homolog in hereditary nonpolyposis colon cancer. Science **263:**1625–1629.
20. **Sakumi, K., M. Foruichi, T. Tsuzuki, T. Kakuma, S. Kawabata, H. Maki, and M. Sekiguchi.** 1993. Cloning and expression of cDNA for a human enzyme that hydrolyzes 8-oxodGTP, a mutagenic substrate for DNA synthesis. J. Biol. Chem. **268:**23524–23530.
21. **Shapiro, M. B., and P. Senapathy.** 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res. **15:**7155–7174.
22. **Shibutani, S., M. Takeshita, and A. P. Grollman.** 1991. Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. Nature (London) **349:**431–434.