Cloning and Expression of the cDNA Encoding the Human Homologue of the DNA Repair Enzyme, *Escherichia coli* Endonuclease III*

(Received for publication, September 27, 1996, and in revised form, December 17, 1996)

Timothy P. Hilbert†, Wenren Chaung‡, Robert J. Boorstein‡, Richard P. Cunningham§, and George W. Teebor‡¶

From the †Department of Pathology and Kaplan Comprehensive Cancer Center, New York University Medical Center, New York, New York 10016 and the §Department of Biological Sciences, SUNY, Albany, New York 12222

We previously purified a bovine pyrimidine hydrate-thymine glycol DNA glycosylase/AP lyase. The amino acid sequence of tryptic bovine peptides was homologous to *Escherichia coli* endonuclease III, theoretical proteins of *Saccharomyces cerevisiae* and *Caenorhabditis elegans*, and the translated sequences of rat and human 3'-expressed sequence tags (3'-ESTs) (Hilbert, T. P., Boorstein, R. J., Kung, H. C., Bolton, P. H., Xing, D., Cunningham, R. P., Teebor, G. W. (1996) *Biochemistry* 35, 2505–2511). Now the human 3'-EST was used to isolate the cDNA clone encoding the human enzyme, which, when expressed as a GST-fusion protein, demonstrated thymine glycol-DNA glycosylase activity and, after incubation with NaCNBH₃, became irreversibly cross-linked thymine glycol-DNA glycosylase activity and, after incubation with NaCNBH₃, became irreversibly cross-linked to a thymine glycol-containing oligodeoxynucleotide, a reaction characteristic of DNA glycosylase/AP lyases. Amino acids within the active site, DNA binding domains, and [4Fe-4S] cluster of endonuclease III are conserved in the human enzyme. The gene for the human enzyme was localized to chromosome 16p13.2–3. Genomic sequences encoding putative endonuclease III homologues are present in bacteria, archaean, and eukaryotes. The ubiquitous distribution of endonuclease III-like proteins suggests that the 5,6-double bond of pyrimidines is subject to oxidation, reduction, and/or hydration in the DNA of organisms of all biologic domains and that the resulting modified pyrimidines are deleterious to the organism.

When a pyrimidine residue in cellular DNA becomes modified by oxidation, reduction, or hydration of its 5,6-double bond, repair is initiated by a DNA-glycosylase activity that cleaves the N-glycosyl bond of the damaged residue, releasing the modified base and creating an abasic (AP) site in the DNA backbone. Such DNA glycosylase activities have been identified in bacteria, yeast, and mammalian species (1–5). The first such enzyme described was *Escherichia coli* endonuclease III, which was identified not on the basis of its DNA glycosylase activity, but rather because it nicked UV-irradiated DNA (9). For this reason it was termed an endonuclease, because it was thought that nicking resulted from enzyme-catalyzed hydrolysis of internucleotide phosphodiester bonds at sites of DNA damage. It has since been determined that the enzyme nicks DNA not via hydrolysis, but by catalyzing β-elimination of the 3'-phosphate group at the AP site formed as a result of the enzyme’s DNA glycosylase activity (10–12). The modified base that was enzymatically released from UV-irradiated DNA proved to be cytosine and/or uracil hydrate (8). Enzymes that effect base release together with strand cleavage via β-elimination are now termed DNA glycosylase/AP lyases and, in addition to endonuclease III, include the Fpg protein of *E. coli* (13), the OGG1 protein of *Saccharomyces cerevisiae* (14, 15), and T4 endonuclease V (16).

DNA glycosylase/AP lyases function through N-acylimine (Schiff’s base) enzyme-substrate intermediates (17). Such enzyme-substrate intermediates can be chemically reduced to stable secondary amines, resulting in irreversible cross-linking of the enzymes to their particular substrates (13, 16–18). We previously used this cross-linking reaction to definitively identify a pyrimidine hydrate-thymine glycol DNA glycosylase/AP lyase purified from calf thymus. Incubation, done under reducing conditions, of a 32P-labeled oligodeoxynucleotide containing a single thymine glycol (5,6-dihydroxy-5,6-dihydrothymine) residue with a 5000-fold purified enzyme preparation resulted in cross-linking of a predominant 31-kDa protein to the oligodeoxynucleotide as determined by SDS-PAGE analysis and phosphor imaging. Tryptic digestion of this protein, followed by microsequencing of several of the resulting peptides demonstrated that the bovine enzyme was homologous to theoretical proteins translated from the genomic DNA of *S. cerevisiae* and *Caenorhabditis elegans*. Both of these theoretical proteins, in turn, were homologues of *E. coli* endonuclease III. The bovine peptide amino acid sequences were also homologous to the translated sequences of 3'-ESTs from *H. sapiens* brain tissue (accession number F04657) and *Rattus sp.* PC 12 cells (accession number H33255) (18). In the current study, we used probes based upon the homologous human 3'-EST, to isolate clones that encode the human homologue of *E. coli* endonuclease III from a splenic cDNA library. Once determined, the cDNA sequence was used to express the enzyme as a functional

---

* This work was supported by National Institutes of Health Grants CA 16669 and 49689 (to G. W. T.), CA 16087 (to Kaplan Cancer Center), GM 07308 (Medical Scientist Training Program for T. P. H.), U. S. Army Medical Research and Development Command DAMD 17–94–J–4416 (to R. J. B.), and GM 46312 (to R. P. C.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† The abbreviations used are: PAGE, polyacrylamide gel electrophoresis; EST, expressed sequence tag; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; IPTG, isopropyl-1-thio-β-D-galactopyranoside; GST, glutathione S-transferase; ORF, open reading frame; HhH, helix-hairpin-helix; FISH, fluorescence in situ hybridization.
recombinant protein and to determine the chromosomal localization of the human gene.

**EXPERIMENTAL PROCEDURES**

Radionucleotides—[α-32P]dCTP (3000 Ci/mmol), [γ-32P]dATP (3000 Ci/mmol), and methyl-[3H]TTP (70–90 Ci/mmol) were purchased from DuPont NEN.

Cloning of the cDNA—Oligodeoxynucleotides based upon the human 3'-EST sequence (accession number F04657) were used to isolate homologous clones from a Superscript human spleen cDNA library in the pCMV-SPORT plasmid vector (Life Technologies, Inc.) using the GENETRAPPER cDNA positive selection system (Life Technologies), according to the manufacturer’s protocol. Briefly, the amplified double-stranded cDNA library was made single-stranded by treatment with the Gene II product (phage P1) endonuclease and E. coli exonuclease III and then hybridized to a biotinylated sense strand-specific oligodeoxynucleotide, P1 (5'-CTGGCACGAGATTCGCTTC). The cDNA-oligodeoxynucleotide hybrids were captured using streptavidin paramagnetic beads. Nonspecifically bound cDNAs were washed away at high stringency, and specifically bound cDNAs were eluted from the paramagnetic beads by denaturing the cDNA-oligodeoxynucleotide hybrids. Selected cDNA clones were then made double-stranded via re-pairing of clones from a second sequence-specific oligodeoxynucleotide, P2 (5'-ATCATGGGACCTCGGTTGG). The selected repaired plasmids were electroverted into the E. coli strain DH5α and plated onto Lennox L agar plates containing 50 μg/ml ampicillin (LB/amp agar).

After 20 h of incubation at 37°C, colonies were analyzed for the presence of the desired cDNA insert under colony PCR, according to the manufacturer’s protocol, using a second set of 3'-EST-specific primers (P3, 5'-CAACAGCGGCTGGCTTTGAACGC; P4, 5'-GTGGGGCTTCGGCCAGCAGCTGT) to maximize specificity of the selection procedure. PCR was conducted as follows: 1 cycle of 95°C for 2 min and 37 cycles of 94°C for 1 min, 60°C for 1 min, 72°C for 1 min, followed by a final cycle of 10 min at 72°C. PCR products were then analyzed by electrophoresis in a 1.2% agarose gel. Colonies that proved positive were selected via sequencing. The sequences of the sense strand, respectively. PCR was conducted as follows: 1 cycle of 95°C for 2 min and 35 cycles of 94°C for 1 min, 65°C for 1 min, and 72°C for 2 min, followed by a final cycle of 10 min at 72°C. The resulting PCR product was digested with BamHI and EcoRI, gel-purified, ligated into gel-purified pGEX-2T vector (Pharmacia Biotech Inc.) that had previously been digested with BamHI and EcoRI, and electroporated into the E. coli strain NB42. Colonies were selected via growth on LB/amp plates, and the presence of the appropriate insert was verified via colony PCR as described above, using primers P3 and P4. Expression of the full-length fusion protein was confirmed via the induction of log phase (A600 = 0.6) 5-ml LB/amp broth cultures with 0.1 mM IPTG for 4 h at 37°C. To prepare total cell SDS lysates, 1-ml aliquots of induced and uninduced cultures were centrifuged at 5000 x g for 2 min, the super- natant was discarded, and the pelleted bacteria were resuspended in 100 μl of SDS-PAGE loading buffer and heated at 95°C for 5 min. Thirty μl of each sample was then analyzed on a 15% Tricine gel. After the gels were stained with Coomassie Blue, induced and uninduced samples were compared to demonstrate the expression of the full-length (65-kDa) fusion protein. Bacterial lysates produced in an identical manner were also run on the SDS-PAGE gel in Fig. 3 in order to demonstrate induction of the GST fusion protein.

Protein Expression and Purification—600 ml of LB/amp broth were inoculated with 10 ml of overnight cultures. Bacteria were grown at 37°C until the A600 reached 0.6. Expression of the fusion protein was induced by incubation with 0.1 mM IPTG for 5 h at 30°C (the lower temperature was used to increase the solubility of the fusion protein). Bacteria were then placed on a shaker at 37°C and 200 rpm, and induced cultures were centrifuged at 3200 x g in 250-ml centrifuge tubes (Corning) for 10 min. The supernatant was discarded, and the pellet was resuspended in 20 ml of sonication buffer (50 mM Tris, pH 8.0, 500 mM NaCl, 5 mM EDTA, 0.5% Triton X-100, 0.25 mM phenylmethylsulfonyl fluoride, 0.1 mg/ml aprotinin). The bacteria were transferred to a 30-ml Corex centrifuge tube and sonicated for 2 min at 70% power using a Heat Systems model 400 ultrasonicator equipped with a model 419 standard tapered microtip. The sonicate was then centrifuged for 15 min at 10,000 x g, and the supernatant was transferred to a 50-ml plastic centrifuge tube containing 1.2 ml of glutathione-agarose 4B affinity medium (volume of medium was measured as a slurry in 20% ethanol, as supplied by the manufacturer) prewashed with 2 x 40 ml of wash buffer (50 mM Tris, pH 8.0, 500 mM NaCl, 5 mM EDTA, 0.5% Triton X-100). The sample was incubated on ice with agitation for 30 min to allow adsorption of the fusion protein. The affinity medium was then pelleted by centrifugation for 2 min at 950 x g. The supernatant was removed by pipetting, and the affinity medium was washed once with 20 ml of wash buffer and 4 times with 40 ml of wash buffer by thorough resuspension of the beads in the appropriate buffer followed by centrifugation at 950 x g for 1 min. After the final wash, the affinity medium was resuspended in 1 ml of wash buffer, transferred to a 2-ml plastic tube, and centrifuged again at 950 x g for 1 min to pellet the beads. The supernatant was removed, and the beads were resuspended in 1 ml of glutathione-agarose elution buffer (100 mM Tris, pH 8.0, 500 mM NaCl, 2.5 mM EDTA, 0.1% Triton X-100, 20 mM glutathione (Sigma)) and incubated for 12 h on ice with agitation. Beads were then quickly pelleted by centrifugation at 950 x g, and the supernatant from the eluted fusion protein was transferred to a fresh tube. All purification procedures from sonication through elution of the fusion protein were carried out at 4°C. The purification yielded 9.9 mg of fusion protein.

As a control, the 26-kDa glutathione S-transferase (GST) of Schistosoma japonicum was expressed from the pGEX-2T vector (without a fusion insert) in the bacterial strain NB42 according to the same pro-
Hybridization overnight at 42°C with 2 M Tris, pH 8.0, 500 mM NaCl, 2.5 mM EDTA, 0.1% Triton X-100, 20 mM glutathione in a quartz cuvette. The optical absorption spectra of the GST fusion protein and the unfolded GST protein were recorded between 200 and 700 nm using a Spectronic Genesys 5 spectrophotometer (Milton Roy). In order to allow comparison of the absorption spectra of the purified GST fusion protein and the previously described, radiation-damaged GST fusion protein (see Fig. 6), the purified proteins were diluted prior to analysis with glutathione-agarose elution buffer to the same absolute protein concentration (5.5 mg/ml).

**FISH Analysis—**FISH Analysis was performed by SeeDNA Biotech Inc. (Dept. of Biology, York University, Ontario, Canada). Lymphocytes isolated from human blood were cultured in an alpha-minimal essential medium supplemented with 10% fetal calf serum and phytohemagglutinin at 37 °C for 68–72 h. The lymphocyte cultures were treated with Br-dUrd (0.18 mg/ml; Sigma) to synchronize the cell population. The synchronized cells were washed 3 times with serum-free medium to release the block and recultured at 37 °C for 6 h in minimal essential medium with thymidine (2.5 mg/ml; Sigma). Cells were harvested and slides were made by using standard procedures, including hypotonic treatment, fixing, and air drying.

To produce a probe for FISH analysis, a 1.1-kilobase pair fragment containing the entire cDNA sequence was excised from an isolated cDNA clone using EcoRI and HindIII, purified, and labeled with biotin-14-dATP using the BioNick labeling kit (Life Technologies) (20). The procedure for FISH analysis was performed according to the previously reported procedures of Heng et al. (21, 22). Briefly, slides were baked at 55 °C for 1 h. After RNase treatment, the slides were denatured in 70% formamide, 2× SSC for 2 min at 70 °C followed by dehydration with ethanol. Probes were denatured at 75 °C for 5 min in a hybridization solution containing 50% formamide, 10% dextran sulfate, and human Cot-restricted DNA. Probes were loaded on the denatured chromosomal slides. After overnight hybridization, slides were washed and analyzed. FISH signals and the DAPI banding pattern were recorded separately by taking photographs. Chromosomal localization was achieved by superimposing FISH signals with DAPI-banded chromosomes (22).

**Northern Blot Analysis—**Two μg of mRNA, isolated from 293T cells using the FastTrack 2.0 mRNA isolation system (Invitrogen), 1 μg of human splenic Poly(A)+ RNA (Clontech), and 5 μg of ϕ324–9.5-kilobase pair RNA ladder (Life Technologies) were electrophoresed on an 11-cm 1.0% agarose-formaldehyde gel. The gel was rinsed with deionized water, and RNA was transferred to a Nytran membrane (Schleicher&Schuell) using the Turboblotter rapid downward transfer system (Schleicher & Schuell), according to the manufacturer’s specifications. Following transfer, the membrane was washed in 2× SSC followed by 1× SSC at 80 °C on a fresh sheet of filter paper, and baked at 80 °C for 1 h. The portion of the membrane that contained the molecular weight markers was cut away and stained with treatment with 5% acetic acid for 15 min and 0.5 M sodium acetate, pH 5.2, with 0.04% methylene blue for 10 min, followed by destaining with water. The baked filter was incubated in prehybridization solution (in 50% formamide, 3× SSC, 0.1 M Tris, pH 7.4, 5 μg ‘Denhardt’s solution’) for 4 h, followed by hybridization overnight at 42 °C with 2× 10^6 cpm of radiolabeled probe/ml of hybridization solution (50% formamide, 3× SSC, 0.1 M Tris, pH 7.4, 5 μg ‘Denhardt’s solution’, 10% dextran sulfate). Following hybridization, the membrane was washed three times for 30 min at 50 °C, successively with 1× SSC, 0.1% SDS, 0.5× SSC, 0.1% SDS, and 0.1× SSC, 0.1% SDS. The membrane was exposed to x-ray film for 24 h at °C. The autoradiogram was matched to the prestained markers to determine the size of the native mRNA. Before hybridization with the cDNA-specific probe, the Northern blot membrane was analyzed by hybridization to a β-actin-specific probe to confirm the integrity of the mRNA. After hybridization to the β-actin probe detected an mRNA species of the predicted size (approximately 2.1 kilobase pairs), the membrane was stripped by boiling for 30 min in 0.1× SSC, 0.5% SDS and 1% SDS, and hybridized to an identical probe with the probe specific for the human homologue of endonuclease III (Fig. 1).

**Preparation of Probes for Northern Blot Analysis—**The β-actin probe was produced by PCR with sequence-specific primers (Clontech) against cDNA made from the RNA of cells taken from a sample of a human bone marrow aspirate. PCR was conducted as follows: 1 cycle of 95 °C for 2 min and 35 cycles of 94 °C for 1 min, 60 °C for 1 min, 72 °C for 1 min, followed by a final cycle of 10 min at 72 °C. The probe was then radiolabeled using the Random Primed DNA Labeling kit (Boehringer Mannheim) and [α-32P]dCTP, and it was purified using Nick-Spin columns (Pharmacia). The specific probe for the human homologue of endonuclease III was prepared by excising the full-length cloning sequence shown in Fig. 1, excising the 2 g of purified plasmid DNA via restriction with EcoRI and BamHI followed by gel purification of the restriction fragment. The probe was radiolabeled and hybridized to the Northern blot membrane as described.

**DNA Glycosylase Assay—**Poly(dA–[3H]dT) was produced by nick translation of the alternating copolymer poly(dA-dT) (Pharmacia) with [γ-32P]ATP, followed by oxidation with osmium tetroxide to form thymine glycol residues (23). Thymine glycol containing poly(dA–[3H]dT) produced in this manner had a specific activity of approximately 1.4 × 10^6 dpm/μg. Thymine glycol DNA-glycosylase assays were carried out against oxidized DNA and the released radioactive product proven to be thymine glycol by high pressure liquid chromatography analysis as described previously (23).

**Sodium Cyanoborohydride-mediated Cross-linking of Fusion Protein to a Thymine Glycol-Containing Oligodeoxynucleotide—**A double-stranded oligodeoxynucleotide containing a single thymine glycol-residue was prepared as described previously (18, 24). The thymine glycol-containing strand was 5′-end-labeled with [γ-32P]dATP, using T4 kinase (Life Technologies) according to the manufacturer’s recommendations, and purified using a ChromaSpin-10 column (Clontech). The purified GST fusion protein, the nonfusion GST protein, and E. coli endonuclease III were reacted with the substrate double-stranded oligodeoxynucleotide in a total volume of 50 μl under the following reaction conditions: 37.3 mM NaCNBH3, 20 mM HEPES, pH 7.5, 46.5 mM KCl, 5 mM EDTA, a 4.0 μM concentration of each oligodeoxynucleotide, and 40 ng/μl protein. In the case of E. coli endonuclease III, this represented approximately a 4-fold molar excess of substrate deoxyoligonucleotide to enzyme. After incubation at room temperature for 2 h, a 25-μl volume of 3× SDS-PAGE loading buffer was added to each sample. Samples were then heated to 90 °C for 5 min and separated by electrophoresis on a 15% Tricine-SDS gel. Following electrophoresis, the gel was stained with Coomassie Blue, wrapped in plastic, and analyzed via autoradiography.

**Gel Electrophoresis—**Prior to electrophoresis all samples were incubated at 95 °C for 5 min in standard SDS-PAGE loading buffer. Fifteen percent Tricine gels (25) were prepared and run using the Mini-Protein II electrophoresis system (Bio-Rad). Gels were run at 90 V for approximately 5 h, completion being determined by the progress of pre-stained low molecular weight electrophoresis standards (Bio-Rad). Gels were then stained with Coomassie Blue.

**RESULTS**

Fig. 1 presents the nucleotide sequence of a cDNA of 1045 base pairs, which contains a putative open reading frame (ORF) of 912 base pairs. This ORF encodes a protein of 304 amino acids with a calculated molecular mass of 33,569 and a calculated pi of 9.85, which is the human homologue of E. coli endonuclease III. The nucleotide sequence data presented in Fig. 1 were obtained from two sources. The sequence of nucleotides 6–1045 was obtained by analysis of clones isolated from a cDNA library, using probes based upon the sequence of the previously described human 3′-EST. The sequence of nucleotides 1–5 was obtained by sequencing the products of 5′-RACE, performed using gene-specific primers based upon the sequence of the longest cDNA clones.

Previously we reported the sequence of four peptides obtained by proteolysis of a purified bovine pyrimidine hydratethymine glycol DNA glycosylase/AP lyase (18). The sequences of those four peptides as well as that of one additional peptide (GEGGEAEHLQAP) derived from the same purified protein are also included in Fig. 1, aligned with the homologous sequences encoded within the ORF of the human cDNA.

The 1045-base pair sequence of Fig. 1 probably represents most, if not all, of the entire full-length cDNA. The Northern blot analysis (Fig. 2) of human splenic and 293T cell (human) mRNA each demonstrate a predominant mRNA species of approximately 1.1–1.2 kilobase pairs, which hybridized to a 32P-labeled probe containing the entire sequence of the ORF. The
difference of approximately 50–150 nucleotides in length between the cDNA sequence presented in Fig. 1 and the native mRNA can be explained by the expected presence of a poly(A) tail of approximately the same length on the native species and perhaps a few more nucleotides 5′ to the first AUG codon.

Fig. 2, lane 3, which contains mRNA extracted from 293T cells, shows a second faint band of higher Mr. Although we think this band is nonspecific, we cannot fully exclude the possibility that it represents mRNA encoding a protein similar to human endonuclease III. Such a situation is present in *S. cerevisiae*, which contains two homologues of *E. coli* endonuclease III, one of which is thought to be nuclear, the second mitochondrial (see "Discussion" and Fig. 8).

To demonstrate that the cDNA sequence of Fig. 1 encoded a functional homologue of endonuclease III, a GST fusion protein was constructed consisting of amino acid residues 8–304 of the ORF fused to the C terminus of the 30-kDa GST protein. SDS-PAGE analysis of the IPTG-induced, affinity-purified fusion protein (Fig. 3) revealed a predominant 65-kDa full-length protein. Two additional lower molecular weight protein species were present in the purified preparation. We believe these to be fragments of the 65-kDa protein that arose through abortive synthesis of the full-length protein or proteolysis occurring before, during, and possibly after cell lysis and affinity purification, due to the action of contaminating cellular proteases.

As demonstrated previously, *E. coli* endonuclease III can be specifically, irreversibly cross-linked to a thymine glycol-containing oligodeoxynucleotide via the reductive stabilization of its characteristic enzyme-substrate intermediate (18). To further confirm that the ORF presented in Fig. 1 encoded a fully functional homologue of *E. coli* endonuclease III, the cross-linking reaction, as described under "Experimental Procedures," was applied to the purified GST-fusion protein. The results of this reaction are illustrated in Fig. 4. When aliquots of the purified GST fusion protein that had been incubated with a 32P-labeled thymine glycol-containing oligodeoxynucleotide...
in the absence (lane 6) or presence (lane 7) of sodium cyanoborohydrate (NaCNBH₃) were compared by SDS-PAGE analysis, it became evident that a portion of the protein had been irreversibly cross-linked to the oligodeoxynucleotide. This is manifested by an increase in the apparent molecular weight of the enzyme resulting in formation of the doublet shown in lane 7. This shift is analogous to that observed when endonuclease III was subjected to the same reductive cross-linking reaction (lane 3) and compared with native endonuclease III (lane 2). No shift of the major protein species was observed when the non-fusion GST protein (lane 3) was incubated under reducing conditions with the thymine glycol-containing oligodeoxynucleotide (lane 4).

An autoradiogram of the gel in Fig. 4A is presented in Fig. 4B. As described, the thymine glycol-containing oligodeoxynucleotide had been 5'-end-labeled with ³²P prior to incubation with the proteins. Thus, cross-linking was confirmed by this autoradiogram in which predominant radioactive species are present only in lanes 2 (E. coli endonuclease III plus NaCNBH₃) and 7 (GST fusion plus NaCNBH₃), which correspond in apparent Mₗ to the shifted species seen on the Coomassie Blue-stained gel. Also evident on the autoradiogram in lane 7 are two visible, less intense, lower molecular weight bands that correspond in position to presumed degradation products of the fusion protein present even after affinity purification (Fig. 3). Presumably these represent cross-linked, partially degraded fusion protein.

After purification, the fusion protein was also analyzed for thymine glycol-DNA glycosylase activity. Fig. 5 presents the V versus [Eₗ] plot in which thymine glycol release is expressed as a function of increasing content of fusion protein. The release of thymine glycol is linear with respect to fusion protein concentration over the amount of protein used. Based on the results of this plot, the specific enzymatic activity of the fusion protein was calculated to be about 1–2% that of genetically engineered E. coli endonuclease III using the same assay (latter assay data not shown). This reduced level of activity is apparently quite common among GST fusion proteins.⁷ GST protein that contained no C-terminal fusion was induced and purified in a manner identical to the fusion protein and assayed for enzymatic activity. This non-fusion GST protein did not demonstrate detectable thymine glycol-DNA glycosylase activity at a protein concentration 3 orders of magnitude higher than that

⁷ R. Schneider, personal communication.
at which the fusion protein was assayed.

As documented previously, *E. coli* endonuclease III contains an iron-sulfur cluster in which a cubane \([4\text{Fe}-4\text{S}]\) moiety is liganded by four cysteine residues. This domain produces a distinctive absorbance at 410 nm (26). Conservation of this \([4\text{Fe}-4\text{S}]\) cluster in the human enzyme was inferred on the basis of the cDNA sequence of Fig. 1, since the putative ORF contains the appropriate four cysteine residues at amino acid positions 282, 289, 292, and 300, and confirmed by taking an absorption spectrum of the purified GST-fusion protein, which revealed that it too absorbed strongly at 410 nm (Fig. 6).

Although purified *E. coli* endonuclease III has a characteristic absorption peak at 410 nm and might be expected to appear blue in solution, the color of solutions containing approximately 0.5 mg/ml or greater of purified endonuclease III are typically yellow-brown (19). Similarly, a solution of the purified GST fusion protein at similar concentrations of protein was also yellow, while a solution of the simultaneously purified non-fusion GST protein was colorless.

In order to determine the chromosomal localization of the gene encoding the mammalian enzyme, FISH analysis was performed as described under "Experimental Procedures." Under the conditions used, hybridization efficiency for our probe was approximately 70% (i.e. among 100 mitotic spreads analyzed, 70 demonstrated binding of the probe to one pair of chromosomes). DAPI banding was used to identify the chromosome pair to which the probe had bound (chromosome 16). The precise localization of the gene (16p13.2) was determined by the summary analysis of 10 pairs of photographs in which the probe signal was matched with the results of DAP1 banding (Fig. 7). There was no additional locus detected by FISH analysis. These results taken together with the presence of a single mRNA species on Northern analysis indicates that the gene for human endonuclease III is a single copy gene.

**DISCUSSION**

The human sequence of Fig. 1 shows a remarkable similarity to that of several other putative homologues of the *E. coli* endonuclease III (National Center for Biotechnology Information (NCBI) sequence ID 119329) found in representative species of all three biologic domains. In bacteria they have been found in both Gram-negative (*Hemophilus influenza*, NCBI sequence ID 1169526) and Gram-positive (*Bacillus subtilis*, NCBI sequence ID 729418) organisms; among archaebacteria, in *Methanococcus jannaschii* (NCBI sequence ID 1510694); and among eukaryotes, in *Schizosaccharomyces pombe* (NCBI sequence ID 1065894), *S. cerevisiae* (NCBI sequence ID 1419843 and 401436), *C. elegans* (NCBI sequence ID 974795), *Rattus sp.* (accession number H33255), and *Homo sapiens* (accession number F04657). The *S. cerevisiae* genome encodes two distinct theoretical homologues of *E. coli* endonuclease III. The alignment of the nine putative homologous sequences using the program Clustal W (version 1.5) (Fig. 8) reveals that a core sequence of amino acids is remarkably well conserved. In bacteria, the core sequence comprises virtually the entire protein. In contrast, the proteins of archaebacteria and eukaryotes have unique extensions at their N and/or C termini. For the sake of clarity, these extensions have been omitted from Fig. 8.

Based upon similarities among several bacterial DNA glycosylases, site-directed mutagenesis studies, and molecular modeling, Thayer et al. (26) identified several regions and residues within the core sequence of amino acids of *E. coli* endonuclease III that could be involved in DNA binding and catalysis. The
regions surrounding glutamine 41 (residue numbers refer to the E. coli endonuclease III amino acid sequence unless otherwise indicated) may form a portion of the substrate binding pocket, in which the damaged pyrimidine fits when in the "flipped out" conformation that the enzyme recognizes. The Helix-hairpin-helix (HhH) motif encoded by the residues surrounding the central LPGVG sequence (residues 114–118) is thought to function in nonspecific DNA recognition. Recently, Doherty et al. (27) have extended this analysis and shown that similar HhH motifs occur in 14 homologous families of DNA-binding proteins, including DNA glycosylases, DNA polymerases, and "flap" endonucleases. Lysine 120 appears to be the nucleophile in the active site of endonuclease III that contributes the ϵ-amino group necessary for the formation of the N-acylimine enzyme-substrate intermediate, characteristic of DNA glycosylase/AP lyases. Aspartic acid 138 has also been implicated as a functional active site residue. All of these residues appear to be well conserved in all of the nine sequences shown. The structure of the E. coli endonuclease III was recently solved (26), and, in light of the high degree of conservation of critical residues, it is likely that the common core sequence of all members of the endonuclease III family will have a similar three-dimensional structure.

In addition to the previously mentioned residues, four highly conserved cysteine residues (187, 194, 197, and 203) have been identified within this common core sequence that contribute to the [4Fe-4S] cluster of E. coli endonuclease III. Examination of the aligned sequences in Fig. 8 reveals that in E. coli endonuclease III and five of its eight putative homologues, including the human enzyme, these four cysteines are arranged according to the consensus sequence Cys-X6-Cys-X2-Cys-X5-Cys. A similar but slightly modified sequence appears in S. pombe (Cys-X6-Cys-X2-Cys-X7-Cys) and M. jannaschii (Cys-X5-Cys-X2-Cys-X7-Cys). Thayer et al. (26) suggested that basic amino acid residues between the first two cysteines of the [4Fe-4S] cluster may form a loop that functions in the nonspecific binding of DNA. While Fig. 8 does not indicate absolute conservation of these residues, some conservation is apparent, especially with respect to arginine 193.

As mentioned previously, the genome of S. cerevisiae encodes two putative homologues of E. coli endonuclease III, one of which (designated Sc non-Fe-S in Fig. 8 (NCBI sequence ID 1419843)) lacks the four-cysteine [4Fe-4S] motif completely and presents an obvious exception to this consensus sequence. However, this sequence also encodes a putative mitochondrial leader sequence (28). Whether pairs of endonuclease III-like proteins, with and without [4Fe-4S] clusters, are present in other eukaryotic organisms and whether the non-Fe-S proteins are mitochondrial remains to be determined.

This interesting question notwithstanding, the presence of endonuclease III-like enzymes in representative species of all three evolutionary domains suggests that the genomic DNA of organisms throughout phylogeny is subject to endogenous stresses that attack the 5,6-double bonds of pyrimidine resi-
of the New York University Skirball Institute Sequencing Facility for DNA sequence.

REFERENCES

1. Brent, T. P. (1973) *Biophys. J.* 13, 399–401
2. Bacchetti, S., and Benne, R. (1975) *Biophys. Acta* 390, 285–297
3. Duker, N. J., and Teebor, G. W. (1975) *Nature* 255, 82–84
4. Ness, I. F., and Nissen-Meyer, J. (1978) *Biochim. Biophys. Acta* 520, 111–121
5. Demple, B., and Linn, S. (1980) *Nature* 287, 203–208
6. Cunningham, R. P., and Weiss, B. (1985) *Proc. Natl. Acad. Sci. U.S.A.* 82, 474–478
7. Doetsch P. W., Helland, D. E., and Haseltine, W. A. (1986) *Biochemistry* 25, 2232–2250
8. Boorstein, R. J., Hilbert, T. P., Cadet, J., Cunningham, R. P., and Teebor, G. W. (1989) *Biochemistry* 28, 6164–6170
9. Radman, M. (1976) *J. Biol. Chem.* 251, 1438–1445
10. Bailly, V., and Verly, W. G. (1987) *Biochem. J.* 242, 565–572
11. Kim, J., and Linn, S. (1989) *J. Biol. Chem.* 264, 2739–2745
12. Manz cudder, A., Gerth, J. A., Absolon, J. M., Stubbe, J., Cunningham, R. P., Wirths, J., and Bolton, P. H. (1991) *Biochemistry* 30, 1119–1126
13. Tsou, J., and Grollman, A. P. (1995) *J. Biol. Chem.* 270, 11671–11677
14. van der Kemp, P. A., Thomas, D., Barbev, R., de Oliveira, R., and Boiteux, S. (1996) *Proc. Natl. Acad. Sci. U.S.A.* 93, 5197–5202
15. Nash, H. M., Bruner, S. D., Scharrer, O. D., Kawate, T., Addona, T. A., Spooner, E., Lane, W. S., and Verdin, G. L. (1996) *Curr. Biol.* 6, 968–980
16. Dods, M. L., Schrock, R. D., III, and Lloyd, R. S. (1993) *Biochemistry* 32, 4294–4296
17. Dods, M. L., Michaels, M. L., and Lloyd, R. S. (1994) *J. Biol. Chem.* 269, 32709–32712
18. Hilbert, T. P., Boorstein, R. J., Kung, H. C., Bolton, P. H., Xing, D., Cunningham, R. P., and Teebor, G. W. (1990) *Biochemistry* 35, 2505–2511
19. Asahara, H., Wistort, P. M., Bank, J. F., Bakerian, R. H., and Cunningham, R. P. (1987) *Biochemistry* 26, 14745–14754
20. Heng, H. H. Q., Squire, J., Tsui, L.-C. (1999) *Proc. Natl. Acad. Sci. U.S.A.* 89, 9509–9513
21. Heng, H. H. Q., and Tsui, L.-C. (1993) *Chromosoma* 102, 325–332
22. Heng, H. H. Q., and Tsui, L.-C. (1996) *Methods in Molecular Biology: In Situ Hybridization Protocols* (Choo, K. H. A., ed) pp. 35–49, Humana Press, Clifton, NJ
23. Higgins, S. A., Frenkel, K., Cummings, A., and Teebor, G. W. (1987) *Biochemistry* 26, 1683–1688
24. Kao, J. Y., Goljer, I., Phan, T. A., and Bolton, P. H. (1993) *J. Biol. Chem.* 268, 17767–17773
25. Shagger, H., and von Jagow, G. (1987) *Anal. Biochem.* 166, 368–379
26. Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P., and Tainer, J. A. (1995) *EMBO J.* 14, 4108–4120
27. Deberty, A. J., Serpell, L. C., and Ponting, C. P. (1996) *Nucleic Acids Res.* 24, 2488–2498
28. Ouellette, B. F., Clark, M. W., Keng T., Storms, R. K., Zhong, W., Zeng, B., Fortin N., Delaney S., Barton, A., Kaback, D. B., and Bussey, H. (1993) *Genome* 36, 32–42
29. Teebor, G. W. (1995) in *DNA Repair Mechanisms and Cancer* (Vos, J.-M. H., ed) pp. 89–108, R. G. Landes Co., Austin, TX
30. Bull, C. J., White, O., Olsen, G. J., Zues, L., Fleischmann, G. G., Sutton, J. A. B., FritzGerald, L. M., and Clayton, R. A. (1996) *Science* 273, 1058–1073
31. Burn, T. C., Connors, T. D., Klinger, K. W., and Landes, G. M. (1995) *Gene* 161, 183–187
32. Sakai, K., Nagahara, H., Abe, K., Obata, H., (1992) *J. Gastroenterol. Hepatol.* 7, 288–292

dues. Previously well characterized substrates of endonuclease III include oxidized pyrimidines such as thymine glycol and 5-hydroxycytosine and hydrates of cytosine and uracil. The oxidation of DNA bases has been primarily attributed to reactive oxygen species formed as byproducts of oxidative metabolism and inflammation. The formation of pyrimidine hydrates has been primarily attributed to the action of UV radiation (reviewed in Ref. 29). The archeon *M. jannaschii* lives beneath the sea and therefore is not exposed to direct sunlight. Furthermore, it is characterized by a reducing rather than an oxidizing metabolism (30). The identification of a homologue of endonuclease III in the genome of this organism suggests that pyrimidines with reduced 5,6-double bonds such as 5,6-dihydrorhodamine may be formed spontaneously in archeon genomic DNA. Perhaps within this evolutionary domain, it is primarily the formation of such reduced rather than oxidized or photohydrated pyrimidine residues that has promoted the conservation of an endonuclease III-like enzyme.

At this time, the specific contribution that the human pyrimidine hydrate-thymine glycol DNA glycosylase/AP lyase activity makes to the maintenance of the genome is uncertain. The human gene encoding this enzyme was localized to the locus 16p13.2-3 by FISH analysis (Fig. 7). The accuracy of this localization was corroborated through the identification of genomic data base nucleotide sequence (accession number L48777) obtained by exon trapping from this same region of chromosome 16 (31), which is 94.1% identical to nucleotides 699–799 of the sequence of Fig. 1. The chromosomal locus of the human endonuclease III homologue is in very close proximity to that of another DNA base excision repair enzyme, 3-methylpurine DNA glycosylase as well as the DNA nucleotide excision gene, ERCC-4. There is no apparent homology among these three proteins, so it seems unlikely that their localization to the same chromosomal region is the result of gene duplication and divergence. Loss of heterozygosity in this region has been reported to occur in 22% of human hepatocellular carcinomas (32). Whether any or all of these DNA repair proteins act as tumor suppressors for human hepatocarcinogenesis remains to be determined.

Acknowledgments—We thank Philip H. Bolton for supplying us with thymine glycol-containing oligodeoxynucleotide, Ajay Chheda for assistance with computer-assisted sequence retrieval, and Archie Cummings, Jr., for technical assistance. We thank Dr. Bernard Goldschmidt for technical assistance. We thank Dr. Bernard Goldschmidt for technical assistance.