Biochemistry

 Volume 35, Number 8 February 27, 1996

Articles

Purification of a Mammalian Homologue of *Escherichia coli* Endonuclease III: Identification of a Bovine Pyrimidine Hydrate-Thymine Glycol DNA-Glycosylase/AP Lyase by Irreversible Cross Linking to a Thymine Glycol-Containing Oligodeoxynucleotide†

Timothy P. Hilbert,‡ Robert J. Boorstein,‡ Hsiang Chuan Kung,§ Philip H. Bolton,§ Dongxia Xing,|| Richard P. Cunningham,⊥ and George W. Teebor*,‡

Department of Pathology and Kaplan Comprehensive Cancer Center, New York University Medical Center, 550 First Avenue, New York City, New York 10016, Department of Chemistry, Wesleyan University, Middletown, Connecticut 06459, and Departments of Chemistry and Biological Sciences, SUNY at Albany, Albany, New York 12222

Received October 20, 1995; Revised Manuscript Received December 20, 1995⊗

ABSTRACT: We purified a homologue of the *Escherichia coli* DNA repair enzyme endonuclease III 5000-fold from calf thymus which, like endonuclease III, demonstrates DNA-glycosylase activity against pyrimidine hydrates and thymine glycol and AP lyase activity (DNA strand cleavage at AP sites via β-elimination). The functional similarity between the enzymes suggested a strategy for definitive identification of the bovine protein based on the nature of its enzyme−substrate (ES) intermediate. Prokaryotic DNA glycosylase/AP lyases function through *N*-acylimine (Schiff's base) ES intermediates which, upon chemical reduction to stable secondary amines, irreversibly cross link the enzyme to oligodeoxynucleotides containing substrate modified bases. We incubated endonuclease III with a ^{32}P-labeled thymine glycol-containing oligodeoxynucleotide in the presence of $NaCNBH_3$. This resulted in an increase in the apparent molecular weight of the enzyme by SDS−PAGE. Phosphorimaging confirmed irreversible cross linking between enzyme and DNA. Identical treatment of the most purified bovine enzyme fraction resulted in irreversible cross linking of the oligodeoxynucleotide to a predominant 31 kDa species. Amino acid analysis of the 31 kDa species revealed homology to the predicted amino acid sequence of a *Caenorhabditis elegans* 27.8 kDa protein which, in turn, has homology to endonuclease III. The translated amino acid sequences of two partial 3′ cDNAs, from *Homo sapiens* and *Rattus sp*., also demonstrate homology to the *C. elegans* and bovine sequences suggesting a homologous family of endonuclease III-like DNA repair enzymes is present throughout phylogeny.

The DNA repair enzyme *Escherichia coli* endonuclease III was initially identified by its nicking activity directed against UV-irradiated DNA (Radman, 1976). Subsequently, it was shown that nicking of UV-irradiated DNA resulted from two enzymatic activities, a DNA-glycosylase which released pyrimidine hydrates from the DNA backbone, yielding an apyrimidinic (AP) site (Boorstein et al., 1989), and an activity which effected strand cleavage via β-elimination of the 3′ phosphate group of the apyrimidinic sugar residue (Bailly & Verly, 1987; Kim & Linn, 1988; Mazumder et al., 1991). The latter activity has been termed an AP lyase to distinguish it from AP endonucleases, such as exonuclease III or endonuclease IV, which catalyze strand cleavage via hydrolysis of phosphodiester bonds (Bailly & Verly, 1989). Endonuclease III is one of a group of enzymes,

† Supported by NIH-CA 16669 and CA 49869 (G.W.T.), CA 16087 (Kaplan Cancer Center), GM 07308 (MSTP for T.P.H.), GM46312 (R.P.C.), USAMRDC DAMD 17-94-J-4416 (R.J.B.), and, in part, DOE-DE-FG02-93ER61598 (P.H.B.).

* Corresponding author.
‡ New York University Medical Center.
§ Wesleyan University.
|| Department of Chemistry, SUNY at Albany.
⊥ Department of Biological Sciences, SUNY at Albany.

⊗ Abstract published in *Advance ACS Abstracts,* February 15, 1996.

0006-2960/96/0435-2505$12.00/0

including T4 endonuclease V and the *E. coli* Fpg protein (MutM), which demonstrate both DNA-glycosylase and AP lyase activities (Demple & Harrison, 1994; Dodson et al., 1994).

In addition to excising pyrimidine hydrates, the DNA-glycosylase activity of endonuclease III also excises pyrimidine glycols, ring-contracted pyrimidine derivatives, such as 5-hydroxymethylhydantoin, and urea residues composed of the N1−C2−N3 atoms of the pyrimidine skeleton (Strniste & Wallace, 1975; Demple & Linn, 1980; Breimer & Lindahl, 1984; Cunningham & Weiss, 1985). Enzyme activities functionally analogous to endonuclease III have been identified in bacteria other than *E. coli*, in yeast, and in mammalian cells and tissues through the use of UV-irradiated, chemically oxidized, and γ-irradiated DNA as substrates (Brent, 1973; Bacchetti & Benne, 1975; Duker & Teebor, 1975; Ness & Nissen-Meyer, 1978; Doetsch et al., 1986). We showed that extracts of Hela cells contained a thymine glycol DNA-glycosylase (Higgins et al., 1987). We also demonstrated that both endonuclease III and Hela cell extracts released cytosine hydrate (as well as its deamination product, uracil hydrate) from UV-irradiated DNA (Boorstein et al., 1989). Kim and Linn (1989) described two, or possibly three, UV-endonuclease activities in Hela cells by monitoring the nicking of UV-irradiated circular DNA. Huq et al. (1992) reported a 25-fold purification of an endonuclease III-like activity from calf thymus and stated that the *N*-terminal sequence of this protein was not homologous to other known proteins.

We undertook the purification of a mammalian endonuclease III-like enzyme from calf thymus by monitoring its DNA glycosylase activity against UV-induced pyrimidine hydrates. The assay measures release of ^{3}H-labeled pyrimidine hydrates and is reproducible and linear with respect to time and protein concentration. The substrate is easily prepared, and, most importantly, the chemical identity of the enzymatically released photoproducts can be corroborated by HPLC analysis. Calf thymus was chosen as the source of enzyme because it contains endonuclease III-like activity and because large amounts of very fresh tissue are available.

A novel approach to the definitive identification of the mammalian enzyme was the application of a chemical reaction which results in the irreversible cross linking of the enzyme to its DNA substrate. *N*-Acylimine (Schiff's base) enzyme−substrate (ES) intermediates are characteristic of the prokaryotic DNA glycosylase/AP lyases described to date. Such intermediates can be irreversibly stabilized through chemical reduction to secondary amines. In such a way T4 endonuclease V (Dodson et al., 1993) and the *E. coli* Fpg protein (Tchou & Grollman, 1995) were irreversibly cross-linked to substrate oligodeoxynucleotides containing a cyclobutane dimer and an 8-oxoguanine residue, respectively. The reductive cross linking of enzyme to an oligodeoxynucleotide permits identification of the mammalian protein by two experimental parameters. The first is an increase in the apparent molecular weight of the enzyme as determined by SDS−PAGE. Second, if the oligodeoxynucleotide is 5′-end-labeled with ^{32}P, the irreversibly cross-linked protein−DNA complex can be detected by autoradiography or phosphorimaging after SDS−PAGE.

On the basis of the results obtained with endonuclease V and the Fpg protein, we anticipated successful irreversible cross linking of *E. coli* endonuclease III to an oligodeoxynucleotide containing one of the enzyme's known substrates, thymine glycol. Assuming that the mammalian enzyme also functions through a *N*-acylimine ES intermediate, we could then apply the reductive cross linking reaction to the purified mammalian enzyme fractions using the same oligodeoxynucleotide. This would permit isolation of the correct protein species from a SDS−polyacrylamide gel in sufficient amount for primary amino acid sequencing.

EXPERIMENTAL PROCEDURES

Buffers. Homogenization buffer: 25 mM HEPES, pH 7.5, 15 mM NaCl, 1 mM DTT, 2 mM EDTA, 0.5 mg/mL leupeptin, 0.7 mg/mL pepstatin, 0.2 mM phenylmethanesulfonyl flouride. HDE: 25 mM HEPES, pH 7.5, 1 mM DTT, 2 mM EDTA.

Enzyme. *E. coli* endonuclease III was purified from *E. coli* strain UC6444 carrying the plasmid pHIT1 as previously described (Asahara et al., 1989).

Radionucleotides. [5,5′-^{3}H]Deoxycytidine 5′-triphosphate (15−30 Ci/mmol) and [*methyl*-^{3}H]thymidine 5′-triphosphate (70−90 Ci/mmol) were purchased from Du Pont/NEN).

Oligodeoxynucleotides. Alternating poly(dG-dC) and poly(dA-dT) were purchased from Pharmacia.

Purification of a Pyrimidine Hydrate DNA-Glycosylase from Calf Thymus. All purification procedures were carried out at 4 °C, unless otherwise indicated. Freshly obtained calf thymus (1.2 kg) was homogenized in a Waring Blendor in 4.8 L of homogenization buffer and further fragmented by sonication in 300 mL aliquots for 3 min at 70% power using a Heat Systems model W-375 sonicator equipped with a model 305 high gain horn. NaCl (4 M) was added to a final concentration of 320 mM and the gelatinous precipitate removed manually by spooling using a 10 mL glass pipette as a stirring rod. The remaining solution was cleared by centrifugation at 10000*g*, filtered through cheesecloth, and diluted with 1.7 volumes of HDE to produce fraction I (4000 mL).

Fraction I was batch extracted with 450 mL (packed volume) of cation exchange resin (SP Fast-flow, Pharmacia) preequilibrated with HDE containing 150 mM NaCl. After the beads settled the supernatant was discarded, and the beads were poured into an XK 26/60 column (Pharmacia). They were washed with 500 mL of HDE containing 150 mM NaCl, followed by a 2 L gradient from 150 to 700 mM NaCl at 4 mL/min. Twenty milliliter column fractions were collected and assayed. Fractions 45−75 were pooled to yield fraction II (620 mL).

Solid ammonium sulfate was added to fraction II, which contained approximately 350 mM NaCl, to a final saturation of 21% (120 g/L solution). The sample was centrifuged at 12000*g* for 20 min to remove precipitate and the supernatant applied to a C 26/40 column (Pharmacia), containing 150 mL (bed volume) of Octylsepharose 4 Fast-flow media (Pharmacia) preequilibrated with HDE, 21% ammonium sulfate, and 300 mM NaCl. The column was washed with 150 mL of HDE, 21% ammonium sulfate, and 300 mM NaCl followed by a 1.5 L gradient beginning with HDE, 21% ammonium sulfate, and 300 mM NaCl and finishing with HDE containing neither ammonium sulfate nor NaCl at 3 mL/min, collected in 20 mL fractions. One milliliter aliquots of the column fractions were dialyzed into HDE and 125 mM NaCl and assayed for enzymatic activity. Active

fractions (31−44) were pooled and dialyzed into HDE and 125 mM NaCl (fraction III, 280 mL).

Fraction III was concentrated by loading onto an HR 10/10 Mono S column (Pharmacia) and eluting via a step increase in NaCl concentration to HDE and 0.5 M NaCl. One milliliter fractions were collected and assayed, and 12 active fractions were pooled. The 12 mL sample was divided into 3 × 4 mL aliquots each of which were fractionated via gel filtration chromatography through a Hiload 26/60, Superdex 75 pg column (Pharmacia), run in HDE and 350 mM NaCl (2.5 mL/min), and collected in 2.5 mL fractions. The gel filtration column was precalibrated with the Gel Filtration Low Molecular Weight Calibration Kit from Pharmacia. Active fractions (70−75, approximately 29 kDa) from each of three column runs were pooled to 45 mL which was diluted from 350 to 125 mM NaCl with 1.8 volumes of HDE. The sample was then loaded onto a HR 5/5 MonoS column (Pharmacia) and concentrated via step elution with HDE and 0.5 M NaCl. Enzymatic activity eluted in six 0.5 mL fractions which were pooled to yield fraction IV (3 mL).

Fraction IV was diluted to 100 mM NaCl with four volumes of HDE, loaded onto a 1 mL single-stranded DNA-cellulose (ssDNA-cellulose, Sigma) HR 5/5 column (Pharmacia) and eluted with a 12.5 mL gradient (100−600 mM NaCl) (0.2 mL/min). Fractions 15−17 were pooled to yield fraction V (1.5 mL).

Preparation of Substrates for DNA-Glycosylase Assays. Poly(dG-[^{3}H]dC) was produced as described previously (Boorstein et al., 1989), by nick translation of poly(dG-dC) (Pharmacia) with [5,5′-^{3}H]dCTP (Du Pont/NEN), and purified using Nick-Spin columns (Pharmacia). Poly(dG-[^{3}H]-dC) produced in this manner had a specific activity of 1.2 × 10^6 cpm/μg. This DNA was then exposed to 400 kJ/m^2 of UV radiation at 254 nm (two 15 W germicidal bulbs) to induce the formation of cytosine hydrate. UV flux was quantitated using a UVX 54 radiometer (UVP Inc., San Gabriel, CA).

Poly(dA-[^{3}H]dT) was produced by the nick-translation of poly(dA-dT) with [*methyl*-^{3}H]dTTP, followed by oxidation of the alternating copolymer with osmium tetroxide to form thymine glycol residues (Higgins et al., 1987). The radiolabeled, oxidized DNA was purified by passing it twice through Nick-Spin columns (Pharmacia). Thymine glycol-containing poly(dA-[^{3}H]dT) produced in this manner had a specific activity of approximately 7 × 10^6 cpm/μg.

DNA-Glycosylase Assays. Pyrimidine hydrate and thymine glycol DNA-glycosylase assays were carried out against UV-irradiated and oxidized DNA substrates, respectively, as follows: enzyme aliquots were incubated with 0.1 μg of substrate DNA in a reaction mixture containing 15 mM HEPES, pH 7.5, 75 mM NaCl, 10 mM EDTA, and 1 mM DTT in a volume of 60 μL for specified periods of time up to 3 h at 37 °C. Reactions were terminated by the addition of 25 μL of 25 mg/mL BSA and 2 mL of acetone, which precipitated both the protein and DNA, leaving in solution only the free modified bases which had been enzymatically cleaved from the DNA backbone. After centrifugation at 8000*g* for 15 min the supernatant was dried, resuspended in water, and analyzed by liquid scintillation counting.

At each step the chemical identity of the released radioactive product was proven to be cytosine by HPLC. The free cytosine hydrate released by the enzyme is unstable, rapidly eliminating water, and is recovered as free cytosine (Boorstein et al., 1989). One unit of enzyme released 1 pmol of cytosine hydrate from 0.1 μg of UV-irradiated poly(dG-[^{3}H]-dC) in 1 min. Enzyme assays lasted from 15 min to 3 h, depending upon the specific activity of the enzyme during the different phases of the purification.

[^{3}H]Thymine glycol released from the oxidized poly(dA-[^{3}H]dT) was identified by HPLC as previously described (Higgins et al., 1987).

AP Nicking Assay. AP-site containing DNA was prepared and nicking activity assayed as described previously (Cunningham & Weiss, 1985). The assay is done in 10 mM EDTA to preclude any Mg^{2+}-dependent AP endonuclease from acting on the substrate.

Preparation of Thymine Glycol-Containing Oligodeoxynucleotide for Cross Linking Studies. Thymine glycol-containing single-stranded oligodeoxynucleotide was prepared as described previously (Kao et al., 1993). The oxidation was carried out on 50 OD_{260} of d(CGCGAT-ACGCC). The complementary 11-mer was synthesized by conventional means.

Cross Linking of Enzyme to Oligodeoxynucleotide. Twenty picomoles of the appropriate oligodeoxynucleotide, either thymine glycol-containing or complementary, was 5′-end-labeled using T4 kinase (Gibco BRL) and [γ-^{32}P]ATP, according to the manufacturer's recommendations, and purified using a Nuc Trap Push Column (Stratagene) preequilibrated in 20 mM HEPES, pH 7.5, 50 mM NaCl, and 5 mM EDTA. The radiolabeled oligodeoxynucleotide was then combined with 200 pmol of nonradioactive oligodeoxynucleotide, and the complementary strand was added at a 1:1 ratio and placed on ice for 30 min. Enzyme was reacted with the substrate double-stranded oligodeoxynucleotide in a total volume of 300 μL under the following reaction conditions: 37.3 mM $NaCNBH_3$, 20 mM HEPES, pH 7.5, 46.5 mM KCl, 5 mM EDTA, 1.5 uM oligodeoxynucleotide, and 15 ng/μL protein. In the case of *E. coli* endonuclease III, this represented a 4-fold molar excess of substrate oligodeoxynucleotide to enzyme. After incubation at 37 °C for 2 h, samples were quick frozen on dry ice, lyophilized, resuspended, and boiled in 35 μL of 1× SDS−PAGE loading buffer 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 phosphorimaging.

Gel Electrophoresis. All samples were lyophilized to dryness and resuspended in standard SDS loading buffer prior to electrophoresis. Fifteen percent Tricine gels were prepared (Shagger et al., 1987) 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 prestained low molecular weight electrophoresis standards (Bio-Rad). Gels were then stained with Coomassie Blue.

Amino Acid Sequence Analysis. Fractions from the ssDNA cellulose column (fraction V) were run on a 15% Tricine-SDS gel and stained with Coomassie Blue. The predominant band, identical to the band which shifted after reductive coupling to the thymine glycol-containing oligodeoxynucleotide, was excised from the gel and sent to the W. M. Keck Foundation microsequencing facility at Yale University, New Haven, CT.

At Yale, the protein was subjected to proteolytic digestion followed by purification on HPLC using a reverse-phase

Table 1: Summary of Purification of a Pyrimidine Hydrate DNA-Glycosylase from Calf Thymus[a]

fraction	total protein (mg)	volume (mL)	total activity (pmol/min)	specific activity [pmol/(min·mg)]	purification (fold)	yield (%)
I	52000	4000	2920	0.056		
II	2420	720	1020	0.421	7.5	34.9
III	264	350	370	1.40	25	12.7
IV	1.3	3.0	36	27.7	495	1.2
V	0.1	1.5	29	290	5180	1.0

[a] Purification steps and fractions are described in the text.

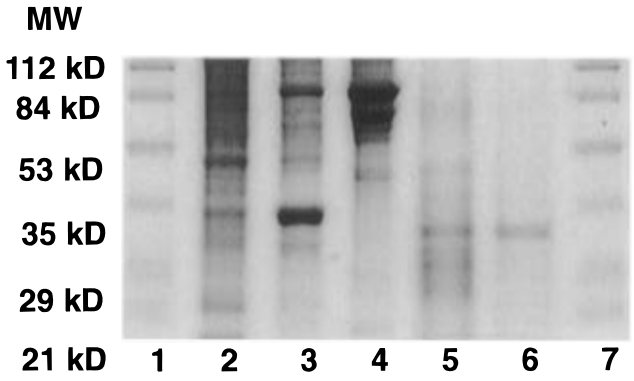

FIGURE 1: SDS−PAGE analysis of the purification fractions. Lanes 1 and 7 contain molecular weight markers. Lanes 2−5 contain fractions I−IV, respectively (described in the text and Table 1). Lane 6 contains material from ssDNA cellulose column fraction 17, which was then pooled with fractions 15 and 16 to yield fraction V.

microbore C18 column. Individual peaks were assayed for purity by laser desorption mass spectroscopy. After a 16 h hydrolysis, amino acid analysis was carried out on a Beckman Model 6300 ion-exchange instrument (Rosenfeld et al., 1992; Elliott et al., 1993; Williams & Stone, 1995; Williams et al., 1995).

The sequence homologies were obtained via the BLAST (Altschul et al., 1990) Network Service of the National Center for Biotechnology Information which accesses the Brookhaven, Swiss, PIR, and GenBank data bases.

RESULTS

Purification of the Mammalian Enzyme. A mammalian homologue of *E. coli* endonuclease III was purified from fresh calf thymus on the basis of its pyrimidine hydrate DNA-glycosylase activity. After the final purification step, ssDNA-cellulose chromatography, the enzyme was purified approximately 5000-fold as estimated by the specific activity of the pyrimidine hydrate DNA-glycosylase and the yield was approximately 1% (Table 1 and Figure 1).

Coelution of DNA-Glycosylase Activities. Successive fractions from the ssDNA-cellulose column were assayed simultaneously for pyrimidine hydrate and thymine glycol DNA-glycosylase activities, both of which have been demonstrated for endonuclease III (Higgins et al., 1987; Boorstein et al., 1989). Figure 2A documents the coelution of the two activities.

Coelution of DNA-Glycosylase and Mg^{2+}-Independent AP Site Nicking Activity. Comparable ssDNA-cellulose purified material from another calf thymus preparation was assayed simultaneously for Mg^{2+}-independent AP-nicking activity and pyrimidine hydrate DNA-glycosylase activity, both of which are also previously documented activities of *E. coli* endonuclease III (Cunningham & Weiss, 1985). The coelution of these two activities is shown in Figure 2B.

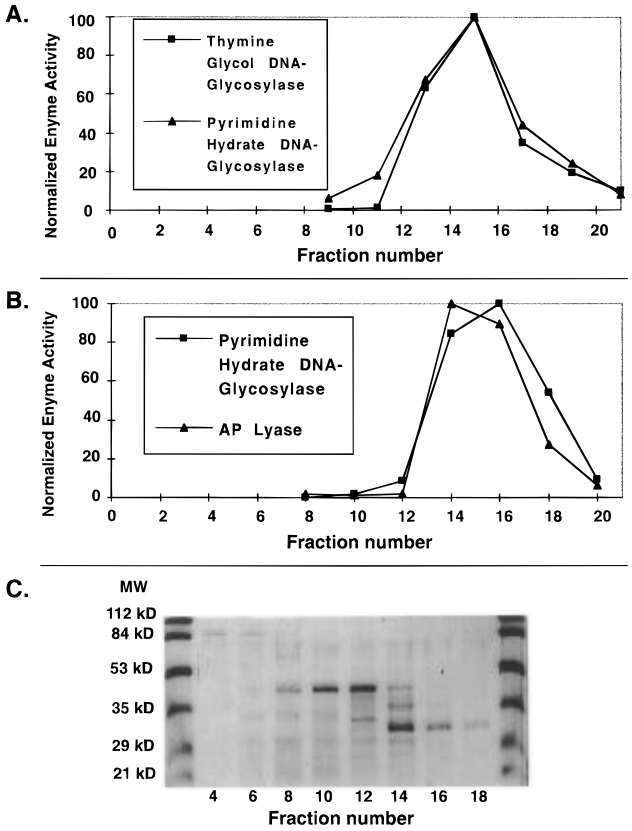

FIGURE 2: (A) Coelution of pyrimidine hydrate and thymine glycol DNA-glycosylase activities from the ssDNA cellulose column. Fractions from the ssDNA cellulose column were assayed simultaneously for both enzyme activities. Activities were normalized by dividing the activity of each fraction by the activity of the fraction with maximum activity. (B) Coelution of pyrimidine hydrate DNA glycosylase activity and AP lyase activity from the ssDNA cellulose column. A second calf thymus preparation was purified through ssDNA cellulose, and elution fractions were analyzed for both enzyme activities, normalized as in Figure 2A. (C) SDS−PAGE analysis of ssDNA cellulose elution fractions (fraction V). A 25 μL aliquot from each of the indicated fractions shown in panel A was analyzed by SDS−PAGE. Fractions 14−18 contain the predominant 31 kDa species. The extreme left and right lanes contain molecular weight markers.

Estimation of the Molecular Weight of the Mammalian Enzyme. The molecular radius of the mammalian DNA-glycosylase, as determined by gel filtration, was approximately 29 kDa. Although ssDNA-cellulose fractions with peak enzymatic activity contained more than one protein species, a predominant band of apparent molecular mass of 31 kDa was present on SDS−PAGE analysis. Moreover, when 25 μL aliquots of successive ssDNA-cellulose column fractions were subjected to electrophoresis and stained with Coomassie Blue, the elution profile of this predominant 31 kDa species, as judged by the intensity of staining, corresponded to that of the two DNA-glycosylase activities (Figure 2C).

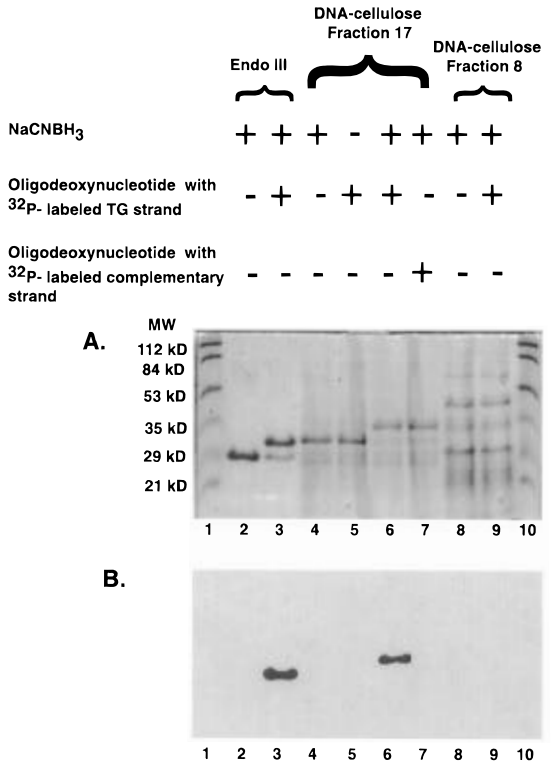

FIGURE 3: (A) SDS−PAGE analysis of *E. coli* endonuclease III and the bovine enzyme after incubation with the thymine glycol-containing oligodeoxynucleotide and $NaCNBH_3$. Lanes 1 and 10 contain molecular weight markers. Lane 2 contains the product of the reaction of *E. coli* endonuclease III and $NaCNBH_3$. Lane 3 contains the product of the same reaction mixture as lane 2 with addition of duplex 5′-^{32}P-labeled oligodeoxynucleotide containing a single thymine glycol (TG) residue. Lane 4 contains the product of fraction 17 eluted from the ssDNA cellulose column and incubated with $NaCNBH_3$ but no oligodeoxynucleotide. Lane 5 contains the product of elution fraction 17 incubated with the 5′-^{32}P-oligodeoxynucleotide but no $NaCNBH_3$. Lane 6 contains the product of elution fraction 17 incubated with both the oligodeoxynucleotide and $NaCNBH_3$. Lane 7 is the same mixture as 6 except that the complementary (non-thymine glycol-containing) oligodeoxynucleotide was 5′-labeled with ^{32}P. Lanes 8 and 9 contain the products of the incubation of ssDNA-cellulose fraction 8, which did not exhibit enzymatic activity, alone or with oligodeoxynucleotide and $NaCNBH_3$, respectively. (B) Phosphorimage of the SDS−PAGE gel of panel A. The lanes are identical to those of panel A.

Reductive Cross Linking of the Enzymes to a Thymine Glycol-Containing DNA Oligodeoxynucleotide. Incubation of purified *E. coli* endonuclease III with duplex DNA (the thymine glycol-containing oligodeoxynucleotide annealed to its complementary strand) in the presence of $NaCNBH_3$ resulted in an increase in the apparent molecular mass of the enzyme as determined by SDS−PAGE (Figure 3A). Lane 2 demonstrates endonuclease III incubated with substrate DNA in the absence of $NaCNBH_3$ and lane 3 in the presence of $NaCNBH_3$. The increase in the apparent molecular mass of the endonuclease III is the result of irreversible cross linking of the enzyme to the oligodeoxynucleotide.

The reductive cross linking reaction was also performed on the most purified preparation of the calf thymus pyrimidine hydrate DNA-glycosylase. A 75 μL aliquot of fraction 17 eluted from the ssDNA cellulose column (Figure 2A) containing purified enzyme of maximal specific activity was incubated with the thymine glycol-containing oligodeoxynucleotide in the presence of $NaCNBH_3$ along with appropriate controls. Lanes 4 and 5 represent ssDNA fraction 17 incubated with $NaCNBH_3$ and no substrate DNA and substrate DNA in the absence of $NaCNBH_3$ respectively. The apparent molecular mass of 31 kDa, as first shown in Figure 2C, did not change under either of these incubations. However, when the reaction mixture contained both substrate DNA and $NaCNBH_3$, the predominant 31 kDa Coomassie Blue-stained band shifted to an apparent molecular mass of 35 kDa, as shown in lanes 6 and 7. As an additional control, fraction 8 eluting from the ssDNA cellulose column (Figure 2A), which contained no enzymatic activity, was also exposed to the conditions of the coupling reaction. Lanes 8 and 9 of Figure 3A contain protein from this fraction incubated with the oligodeoxynucleotide in the presence or absence of $NaCNBH_3$. As can readily be seen, no shift of the visible protein bands occurred.

Additional proof of irreversible cross linking is demonstrated in Figure 3B. The thymine glycol-containing oligodeoxynucleotide had been 5′-end-labeled with ^{32}P prior to the coupling reactions. A phosphorimage of the gel in Figure 3A demonstrated only two bands. Figure 3B shows a single band in lane 3 which corresponds to the position of the shifted cross linked endonuclease III. The single band in lane 6 corresponds to the position of the predominant Coomassie-Blue stained species from calf thymus, which had also shifted after cross linking. Under the denaturing conditions (boiling) used to prepare samples for the SDS gel, the complementary oligodeoxynucleotide strand does not remain associated with the protein. When the complementary, rather then the thymine glycol-containing oligodeoxynucleotide was ^{32}P-labeled, the protein shifted (Figure 3A, lane 7) but did not appear on the phosphorimage (Figure 3B, lane 7). Thus, lanes 6 and 7 of Figure 3B prove that the bovine enzyme cross linked only to the oligodeoxynucleotide containing the thymine glycol residue, thereby confirming that the irreversible cross linking resulting from chemical reduction is exclusively dependent upon formation of a specific ES intermediate. There was no evidence of any binding of oligodeoxynucleotide to the proteins which did not contain enzyme activity, further corroborating that the reductive cross linking reaction was absolutely specific.

Amino Acid Sequence Data. Four peptides derived from a proteolytic digest of the purified bovine protein were sequenced yielding sequences of 14, 15, 22, and 23 amino acids. None of these sequences demonstrated direct similarity to *E. coli* endonuclease III by initial BLAST analysis. However, the 22 amino acid peptide sequence demonstrated considerable similarity to a portion of two predicted full length protein sequences from *Caenorhabditis elegans* (accession no. Z05874) (Wilson et al., 1994) with $P(N) = 0.00053$ and *Saccharomyces cerevesiae* (accession no. L05146) with $P(N) = 0.0063$). Both the *C. elegans* and *S. cerevesiae* proteins, in turn, bear similarity to *E. coli* endonuclease III (accession no. J02857). When compared with the sequence of endonuclease III via BLAST, the *C. elegans* and the *S. cerevesiae* sequence yielded $P(N)$ values of 9.1×10^{-25} and 1.9×10^{-7}, respectively. This same bovine polypeptide demonstrates an even greater degree of similarity to two recently submitted partial 3′ cDNA sequences, from *Homo sapiens* (accession no. F04657) with $P(N) = 6.8 \times 10^{-9}$ and *Rattus sp.* (accession no. H33255) with $P(N) = 1.8 \times 10^{-7}$.

Figure 4 demonstrates the alignment of the *E. coli* endonuclease III amino acid sequence, with the primary

```
A:1 MNKAKRLEILTRLRENNPHPTTELNFSSPFELLIAVLLSAQATDVSVNKATAKLYPVANTPAAMLELGVE 70
B:1  MRKDMIAPVDTMGCHKLADPLAAPPVHRFQVLVALMLSSQTRDEVNAAAMKRLKDHGLSIGKILEFKVP 69
C:          PVDQLGAEHCFDPSA                                        LTVDSILQTDDS

A:71 GVKTYIKTIGLYNSKAENIIKTCRILLEQHNGEVPEDRAALEALPGVGRKTANVVLNTAFGWPT.IAV 137
B:70 DLETILCPVGFYKRKAVYLQKTAKILKDDFSGDIPDSLDGLCALPGVGPKMANLVMQIAWGECVGIAV 137
C:   TLGALIVPVGF                                                 QGTVNGIAV

A:138 DTHIFRVCNRTQFAPGKN.VEQVEEKLLKVVPAEFKVDCHHWLILHGRYTCIARKPRCGSCIIEDLC 203
B:138 DTHVHRIS.NRLGWIKTSTPEKTQKALEILLPKSEWQPINHLLVGFGQMQCQPVRPKCGTCLCRFTC 203
C:    XTHVP                          *LWSEINGLLVGFGQQTCLPIRP
D:                                 WLPR?LWHEINGLLVGFGQQTCLPVHPRCHACLNQALC
E:       HRIANRLKWTKKMTKSPEETRRNLE?WLPRVLWSEINGLLVGFGQ?ICLPVHPRCQACL?KALC

A:204 EYKEKVDI 211
B:204 PSSTAKNVKSETEETSTSIEVKQEVEDEFEDEKPAKKIKKTRKTRTKIEVKTESET 259
C:
D:    PAAQGL
E:    PAAQGL
```

FIGURE 4: Amino acid sequence alignment of *E. coli* endonuclease III (line A), *C. elegans* translated protein (line B), bovine primary amino acid sequences for peptides of 15, 23, 14, and 22 amino acids, respectively (line C), and *H. sapiens* and *Rattus sp.* sequences obtained by translation of partial cDNA sequence (lines D and E, respectively). X in sequence C represents an indeterminate amino acid residue. ? in sequences D and E represents indeterminate nucleotide sequences. The six amino acid region presented in boldface and italic constitutes a portion of the active site of endonuclease III. A 22 amino acid region of near identity between the predicted *C. elegans* sequence, the primary bovine sequence, and the *H. sapiens* and *Rattus sp.* translated partial cDNA sequences is presented in boldface. The four cysteine residues presented in double underlined type represent the ligands of the iron−sulfur cluster of *E. coli* endonuclease III.

amino acid sequences of the bovine polypeptides, and the predicted amino acid sequences of the *C. elegans*, *H. sapiens*, and *Rattus sp.* proteins derived by translation of their respective nucleotide sequences. The boldface sequence marked with the asterisk represents the 22 amino acid bovine polypeptide found to be most similar to the *C. elegans* and the *H. sapiens* and *Rattus sp.* sequences, as determined by the BLAST program using the default Blosum 62 as the algorithmic matrix and the default expected cutoff value of 10. The other peptides were aligned by the BLAST program after we raised the expected cutoff value from 10 to 100.

DISCUSSION

We purified a mammalian pyrimidine hydrate DNA-glycosylase 5000-fold from calf thymus (summarized in Table 1). The most purified fractions, after elution from ssDNA cellulose, also demonstrated thymine glycol DNA-glycosylase and AP lyase activities. The AP nicking assay, performed in the presence of 10 mM EDTA, has previously been shown to specifically correspond to β-elimination (Mazumder et al., 1991). The elution profiles of the two glycosylase activities and the AP lyase activities were superimposable (Figure 2A,B). The fact that these activities coeluted in the most purified calf thymus fractions strongly suggests that they are contained within the same protein.

When identical volumes of successive column fractions were analyzed by SDS−PAGE, there was strong correspondence between the intensity of staining of a predominant 31 kDa band in the active fractions and the enzyme activities, but several other protein species were also present which could have represented the bovine pyrimidine hydrate-thymine glycol DNA-glycosylase/AP lyase.

Therefore, to definitively identify the bovine enzyme, we took advantage of a reductive cross linking reaction which had already been applied to T4 endonuclease V and the *E. coli* Fpg protein. We first demonstrated that, in the presence of $NaCNBH_3$, purified *E. coli* endonuclease III would form a stable cross link to an oligodeoxynucleotide containing one of its substrates, thymine glycol. The apparent increase in molecular weight of the purified enzyme (Figure 3A, lane 3) together with the phosphorimaging data (Figure 3B, lane 3) demonstrated unequivocally that the bacterial enzyme was irreversibly cross linked to the substrate oligodeoxynucleotide. Thus, *E. coli* endonuclease III was cross linked to a substrate DNA oligodeoxynucleotide in a manner analogous to T4 endonuclease V and the *E. coli* Fpg protein, confirming that it also functions via an *N*-acylimine ES intermediate.

We applied the same reaction to the most purified bovine enzyme fraction and showed that only the predominant 31 kDa protein species was irreversibly cross linked to the same thymine glycol-containing oligodeoxynucleotide. The specificity of the reaction was confirmed by separately 5′-end-labeling either the thymine glycol-containing or complementary strand of the substrate double-stranded oligodeoxynucleotide. The increase in the apparent molecular mass of the 31 kDa protein occurred independently of which DNA strand was labeled; however, only when the thymine glycol-containing oligodeoxynucleotide was labeled did a band corresponding to the shifted protein appear on phosphorimage of the gel (Figure 3B). That this strategy enabled us to successfully identify the bovine analog of *E. coli* endonuclease III in a relatively complex mixture of mammalian proteins was contingent on the fact, unproven until now, that the mammalian enzyme and the bacterial enzyme both function through a *N*-acylimine ES intermediate.

The primary amino acid sequence data confirms that the purified 31 kDa protein species we identified by reductive cross linking is a mammalian homologue of endonuclease III. The aligned sequences of Figure 4 demonstrate the homology between the bovine and *C. elegans* proteins extending into the region which constitutes the iron−sulfur cluster of *E. coli* endonuclease III (Thayer et al., 1995). This iron−sulfur cluster motif contains four cysteine residues at endonuclease III positions 187, 194, 197, and 203 and has been shown to be a DNA binding domain. The *H. sapiens* and *Rattus sp*. partial 3′ cDNA sequences also contain four cysteine residues which align with those of *E. coli* endonuclease III and the *C. elegans* sequence. Thus, it seems probable that the *E. coli*, *C. elegans*, and mammalian enzymes all share a common mode of DNA binding. A second bovine peptide and the *C. elegans* predicted protein both align with a region containing a known active site amino

acid of endonuclease III, aspartic acid 138 (Figure 4, bold, italics). Another critical active site residue in *E. coli* endonuclease III is lysine 120 which probably contributes the ϵ-amino group necessary for the formation of the *N*-acylimine ES intermediate (Thayer et al., 1995). Since we have demonstrated such an ES intermediate for the bovine enzyme, it is probable that mammalian DNA glycosylase/AP lyases will prove to have a lysine residue as part of their active sites. In conclusion, given the similarities in amino acid sequence, including active sites and DNA binding domains among the *E. coli* endonuclease III, the purified bovine enzyme and the predicted sequences of the *C. elegans*, *H. sapiens*, and *Rattus sp.* proteins, we suggest that there is a homologous family of endonuclease III-like DNA repair enzymes present throughout phylogeny.

ACKNOWLEDGMENT

We thank Mr. Archie Cummings for invaluable technical assistance during the purification of the bovine enzyme. We also thank Maria Thayer for helpful discussions on protein similarities. We thank Ken Williams and Kathy Stone of the W. M. Keck Foundation Biotechnology Resource Laboratory at Yale University for carrying out the protein sequencing and for helpful discussion.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Meyers, E. W., & Lipman, D. J. (1990) *J. Mol. Biol. 215*, 403−410.
Asahara, H., Wistort, P. M., Bank, J. F., Bakerian, R. H., & Cunningham, R. P. (1989) *Biochemistry 28*, 4444−4449.
Bacchetti, S., & Benne, R. (1975) *Biochim. Biophys. Acta 390*, 285−297.
Bailly, V., & Verly, W. G. (1987) *Biochemical J. 242*, 565−572.
Bailly, V., & Verly, W. G. (1989) *Nucleic Acids Res. 17*, 3617−3618.
Boorstein, R. J., Hilbert, T. P., Cadet, J., Cunningham, R. P., & Teebor, G. W. (1989) *Biochemistry 28*, 6164−6170.
Breimer, L. H., & Lindahl, T. (1984) *J. Biol. Chem. 259*, 5543−5548.
Brent, T. P. (1973) *Biophys. J. 13*, 399−401.
Cunningham, R. P., & Weiss, B. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 474−478.
Demple, B., & Linn, S. (1980) *Nature 287*, 203−208.
Demple, B., & Harrison, L. (1994) *Annu. Rev. Biochem. 63*, 915−948.
Dizdaroglu, M., Laval, J., & Boiteux, S. (1993) *Biochemistry 32*, 12105−12111.
Dodson, M. L., Schrock, R. D., III, & Lloyd, R. S. (1993) *Biochemistry 32*, 8284−8290.
Dodson, M. L., Michaels, M. L., & Lloyd, R. S. (1994) *J. Biol. Chem. 269*, 32709−32712.
Doetsch P. W., Helland, D. E., & Haseltine, W. A. (1986) *Biochemistry 25*, 2212−2220.
Duker, N. J., & Teebor, G. W. (1975) *Nature 255*, 82−84.
Elliott, J. I., Stone, K. L., & Williams, K. R. (1993) *Anal. Biochem. 211*, 94−101.
Higgins, S. A., Frenkel, K., Cummings, A., & Teebor, G. W. (1987) *Biochemistry 26*, 1683−1688.
Huq, I., Haukanes, B. I., & Helland, D. E. (1992) *Eur. J. Biochem. 206*, 833−839.
Kao, J. Y., Goljer, I., Phan, T. A., & Bolton, P. H. (1993) *J. Biol. Chem. 268*, 17787−17793.
Kim, J., & Linn, S. (1989) *J. Biol. Chem. 264*, 2739−2745.
Mazumder, A., Gerlt, J. A., Absolon, J. M., Stubbe, J., Cunningham, R. P., Withka, J., & Bolton, P. H. (1991) *Biochemistry 30*, 1119−1126.
Ness, I. F., & Nissen-Meyer, J. (1978) *Biochim. Biophys. Acta 520*, 111−121.
Radman, M. (1976) *J. Biol. Chem. 251*, 1438−1445.
Rosenfeld, J., Capdeville, J., Guillemot, J., & Ferrara, P. (1992) *Anal. Biochem. 203*, 173−179.
Shagger, H., & von Jagow, G. (1987) *Anal. Biochem. 166*, 368−379.
Strniste, G. F., & Wallace, S. S. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 1997−2001.
Tchou, J., & Grollman, A. P. (1995) *J. Biol. Chem. 270*, 11671−11677.
Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P., & Tainer, J. A. (1995) *EMBO J. 14*, 4108−4120.
Williams, K. R., & Stone, K. L. (1995) in *Techniques in Protein Chemistry VI* (Crabb, J., Ed.) pp 143−152, Academic Press, New York.
Williams, K. R., Samander, S., Stone, K. L., Saylor, M., & Rush, J. (1995) in *Protein Protocol Handbook* (Walker, J., Ed.) Humana Press.
Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., et al. (1994) *Nature 368*, 32−38.

BI952516E