Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 8508–8513, July 1998
Biochemistry

3,N^4-ethenocytosine, a highly mutagenic adduct, is a primary substrate for *Escherichia coli* double-stranded uracil-DNA glycosylase and human mismatch-specific thymine-DNA glycosylase

(ethenoadducts/base excision DNA repair/lipid peroxidation)

MURAT SAPARBAEV AND JACQUES LAVAL†

Groupe Réparation des lésions Radio-et Chimio-Induites, UMR 1772, Centre National de la Recherche Scientifique, Institut Gustave Roussy, 94805 Villejuif Cedex, France

Communicated by Richard B. Setlow, Brookhaven National Laboratory, Upton, NY, May 11, 1998 (received for review January 28, 1998)

ABSTRACT **Exocyclic DNA adducts are generated in cellular DNA by various industrial pollutants such as the carcinogen vinyl chloride and by endogenous products of lipid peroxidation. The etheno derivatives of purine and pyrimidine bases 3,N^4-ethenocytosine (εC), 1,N^6-ethenoadenine (εA), N^2,3-ethenoguanine, and 1,N^2-ethenoguanine cause mutations. The εA residues are excised by the human and the *Escherichia coli* 3-methyladenine-DNA glycosylases (ANPG and AlkA proteins, respectively), but the enzymes repairing εC residues have not yet been described. We have identified two homologous proteins present in human cells and *E. coli* that remove εC residues by a DNA glycosylase activity. The human enzyme is an activity of the mismatch-specific thymine-DNA glycosylase (hTDG). The bacterial enzyme is the double-stranded uracil-DNA glycosylase (dsUDG) that is the homologue of the hTDG. In addition to uracil and εC-DNA glycosylase activity, the dsUDG protein repairs thymine in a G/T mismatch. The fact that εC is recognized and efficiently excised by the *E. coli* dsUDG and hTDG proteins *in vitro* suggests that these enzymes may be responsible for the repair of this mutagenic lesion *in vivo* and be important contributors to genetic stability.**

A number of chemical carcinogens induces the formation of cyclic adducts in DNA. These include industrial chemicals such as vinyl chloride (1) and the widespread environmental compound ethyl carbamate (2, 3). Ethyl carbamate is a carcinogen that is metabolized to vinyl carbamate, then oxidized by cytochrome P450 to the electrophyle vinyl carbamate epoxide that reacts with RNA and DNA bases to form etheno-bridged adducts (4–6). Similarly, two oxidized metabolites of vinyl chloride, 2-chloroacetaldehyde and 2-chloroethylene oxide, predominantly react with purine and pyrimidine residues in DNA and RNA producing 1,N^6-ethenoadenine (εA), 3,N^4-ethenocytosine (εC), N^2,3-ethenoguanine (εG), and 1,N^2-ethenoguanine (1,N^2-εG) (7, 8). Furthermore, the generation of exocyclic DNA adducts by products of membrane lipids peroxidation has been demonstrated (9, 10). The level of εC present in human liver has been found to be 2.8 ± 0.9 per 10^7 bases (11).

The mutagenic potential of cyclic DNA lesions has been established. In *Escherichia coli* and simian kidney cells, εC mostly produces εC·G to A·T transversions and εC·G to T·A transitions (12, 13). For single-stranded shuttle vestor containing a single εC residue the targeted mutation frequency yields a 2% in *E. coli*, 32% in SOS-induced *E. coli* cells, and 81% in simian kidney cells (13). For comparison, the apparent mutation frequency measured in *E. coli* for a single C_8-hydroxyguanine residue in double-stranded M13 phage DNA is 0.3% (14). In mammalian cells, εA residues mainly lead to εA·T to G·C transitions (15). In bacterial systems, εG has miscoding properties producing εG·C to A·T transitions (16), and 1,N^2-εG leads to G·C to A·T transitions (17). The repair of εA and εG adducts by DNA glycosylases present in crude cells extracts has been described (18, 19). The excision of εA by pure 3-methyladenine-DNA glycosylases of various origins have shown that the human enzyme is by far the most efficient (20). The purified AlkA protein releases εG when present in DNA (21).

An enzymatic activity repairing εC has been identified and partially purified from human cells and shown to be different from the ANPG protein (22, 23) but has not been characterized at the molecular level. Surprisingly, until now there was a complete lack of information about the repair of εC in prokaryotes.

In the present study, using an εC-containing duplex oligonucleotide, we purified to homogeneity an εC-DNA glycosylase (εCDG) activity from *E. coli* cell extract. It was identified as the double-stranded uracil-DNA glycosylase (dsUDG) (24). The protein acts with an unusual efficacy. We have identified, by analogy (25, 26), the human mismatch-specific thymine DNA glycosylase (hTDG) (27) as the enzyme excising εC in human cells, also with a good efficiency. These observations suggest a possible role of these proteins *in vivo* to counteract the genotoxic effects of εC residues.

MATERIALS AND METHODS

Oligonucleotides. The 34-mer oligonucleotide 5′-AAATACATCGTCACCTGGG**X**CATGTTGCAGATCC-3′, where at position 20, X = εC, U, or T, was purchased from Genset (Paris). This sequence previously was used to identify the ethenoadenine and the hypoxanthine-DNA glycosylases (20, 28). These sequences will be referred to as (εC-34), (U-34), or (T-34). The 34-mer oligonucleotide containing εC, U, or T was ^{32}P-labeled at the 5′-end by T4 polynucleotide kinase or at the 3′-end by terminal transferase, yielding [^{32}P] 5′-end- or 3′-end-labeled εC-34, U-34, or T-34 oligonucleotides. Four complementary oligonucleotides, containing dA, dG, dC, or T opposite to X were synthesized by E. Lescot (this laboratory). The duplex oligonucleotides, made by annealing (εC-34), (U-34), or (T-34) with the complementary oligonucleotides as already described (28), will be referred to as εC-34/G, εC-34/A, εC-34/T, εC-34/C or U-34/G, U-34/A, U-34/T, U-34/C or T-34/G, T-34/A, T-34/T, T-34/C when the base

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

Abbreviations: εC, 3,N^4-ethenocytosine; εA, 1,N^6-ethenoadenine; εG, N^2,3-ethenoguanine; 1,N^2-εG, 1,N^2-ethenoguanine; εCDG, εC-DNA glycosylase; dsUDG, double-stranded uracil-DNA glycosylase; hTDG, human thymine-DNA glycosylase; AP, apurinic.

†To whom reprint requests should be addressed. e-mail: jlaval@igr.fr.

opposite to the adduct is G, A, T, C, respectively. We also used the following 34-mer oligonucleotide containing thymine at position 19 (T-19): 5′-CGGTATCCACCAGGTCATTAATA-ACGATGAAGCC-3′ annealed to a complementary oligonucleotide containing dG at position 19 (synthesized by E. Lescot of this laboratory). This duplex oligonucleotide will be referred to as T-19/G.

Enzymes. Xth protein, terminal transferase, and molecular biology products were purchased from Boehringer Mannheim. T4 polynucleotide kinase was purchased from New England Biolabs. The purification of the *E. coli* FPG protein (29), Nfo protein (30), UNG, Nth protein (31), Tag, and AlkA protein (32, 33) was performed as described. The ANPG40 (34), ANPG60 (35), and APDG60 (32) proteins were purified to apparent homogeneity from extracts of *E. coli* BH290 (*tag*, *alkA*) harboring plasmids containing the ANPG40, ANPG60, and APDG60 cDNAs, respectively. The activity of the various proteins was tested by using their classical substrates and was checked just before their use.

DNA Glycosylase Assays. The DNA glycosylase assay and analysis of the products of the reaction were performed as described (28) but using εC-34/G as substrate. Incubations were made at 30°C for the human enzyme and at 37°C for the bacterial enzyme, unless otherwise stated. The products released by the enzymes were characterized by HPLC using a Beckman System Gold equipped with a C_{18} μBondapak column. The column was isocraticaly eluted at 1 ml/min with 50 mM $NH_4H_2PO_4$ (pH 4.5) containing 10% methanol (vol/vol). The elution of the products was monitored by UV absorption at 270 nM and 280 nM.

Enzymes Purification. *Purification of the bacterial εCDG.* To purify the εCDG from bacterial cells, we chose the *E. coli* strain RZ1032 (*ung*, *dut*). Cell-free extract was prepared by using freezing-thawing cycles without lysozyme. After overnight culture (1 liter), the cells were harvested by centrifugation and washed with 250 ml of buffer A [0.3 M Tris·HCl, pH 8.0/5 mM EDTA/0.1 mM phenylmethylsulfonyl fluoride (PMSF)]. They were resuspended in the same buffer (1 g cells/8 ml buffer) and stored frozen at −20°C. Frozen cells were thawed in ice-water bath for 1 h and then heated for 10 min at 37°C. Then the mixture was placed in a dry ice-ethanol bath for 15 min. This procedure was repeated 3–4 times, and the lysate was clarified by centrifugation (10,000 *g* for 10 min at 4°C). The supernatant was filtered through 22 μm (fraction I) and made 1.7 M in ammonium sulfate. The resulting precipitate was removed by centrifugation, and the supernatant (fraction II) was applied on a Phenyl-Sepharose (Pharmacia) column (1.5 × 3.5 cm) and rinsed with buffer B (buffer A containing 1.7 M ammonium sulfate). The proteins were eluted by using a linear gradient (total volume 60 ml), 0–100% of buffer C (20 mM Tris·HCl, pH 8.0/2 mM EDTA/2.5 mM β-mercaptoethanol/0.1 mM PMSF/5% glycerol). Fractions containing εCDG activity were pooled and dialyzed against buffer D (as buffer C, but without glycerol). The dialyzed solution (fraction III) was applied on a Mono Q HR 5/5 FPLC column using an FPLC system (Pharmacia). The flow-through (fraction IV) was collected and loaded on a Mono S HR 5/5 column. The column was rinsed with buffer D, and a gradient from 0 to 800 mM NaCl in buffer D (15 ml, 30 min, 0.5 ml/min) was used to develop the column. Fractions containing the εCDG activity were supplemented with glycerol (50%) and stored at −20°C.

Purification of the hTDG protein. The plasmid DNA pT7-hTDG containing the TDG cDNA coding for the hTDG (27) was provided by J. Jiriçny (Institute for Medical Radiobiology, Zurich, Switzerland). The purification procedure was similar to the method described by Neddermann *et al.* (27) but with modifications. A 2-liter culture of *E. coli* BL21 (DE3) containing pT7-hTDG was grown at 30°C to $A_{600} = 0.9$. It then was cooled to 22°C and induced with 1 mM isopropyl β-D-thiogalactoside for 12 h at 22°C. Bacterial cells were collected by centrifugation, washed with TEG buffer (25 mM Tris, pH 8.0/0.1 mM phenylmethylsulfonyl fluoride/2 mM EDTA/10% glycerol/2.5 M β-mercaptoethanol). They were resuspended in 80 ml of the same buffer and stored at −20°C. The cells were lysed by sonication at 4°C, and the lysate was adjusted to 0.1 M NaCl and centrifuged (30,000 rpm for 60 min, in a Beckman 42.1 rotor). The supernatant (80 ml, fraction I) was loaded on a column of DEAE-Sepharose Fast Flow Resin (Pharmacia) (1.44 × 4.6 cm). The G/T mismatch-specific DNA glycosylase activity was recovered in the flow-through fraction whereas most of *E. coli* εCDG activity remained on the QMA column. The flow-through fraction (fraction II) then was applied on a POROS 20 HS (1 ml column) (Boehringer Mannheim) equilibrated with TEG buffer containing 0.1 M NaCl. The human protein was eluted by using a linear gradient of 0.1 to 1 M NaCl in TEG buffer. The fractions containing G/T mismatch-specific DNA glycosylase activity were collected and diluted with TEG buffer to 0.1 M NaCl (fraction III) and loaded on Mono Q HR 5/5 column (Pharmacia). Again, the G/T mismatch-specific DNA glycosylase activity did not stick to the column and was recovered in the flow-through fraction (fraction IV). Fraction IV was directly loaded on a Mono S HR 5/5 column (Pharmacia). The G/T mismatch-specific DNA glycosylase activity was eluted by using a 0.1–0.8 M NaCl linear gradient developed for 30 min at 0.5 ml/min. The hTDG activity peak eluted at about 0.5 M NaCl.

RESULTS

Identification of an Activity Excising εC in Crude Extracts of *E. coli*. Although the εCDG activity has been identified in mammalian cells extracts, such enzyme has never been described in *E. coli* extracts. As shown in Fig. 1, incubation of the 5′-end-labeled oligonucleotide εC-34/G containing a single εC residue at position 20, with crude extracts from *E. coli* AB1157 and subsequent analysis by electrophoresis on a denaturing polyacrylamide gel, shows the appearance of reaction products migrating at the position of a 19-mer and a lower migrating band presumably generated by a nonspecific exonuclease activity present in *E. coli* extracts. When a 3′-end-labeled εC-34/G duplex oligonucleotide is used as substrate, the product of the reaction migrates at the position of a 14-mer. These results suggest that the incision occurs by the action of a DNA glycosylase followed by the cleavage of DNA by an

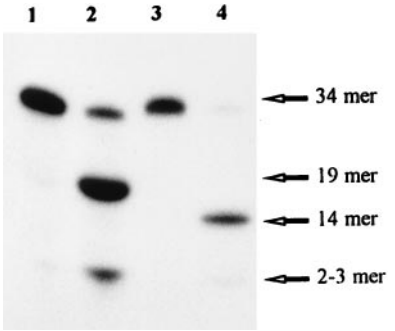

FIG. 1. Cleavage of oligonucleotides containing εC residues by *E. coli* cells extract. 5′-end or 3′-end ^{32}P-labeled oligonucleotides containing εC residues were incubated with *E. coli* AB1157 cell extract. The products of the reaction were separated on a 20% denaturing PAGE. Lane 1, 5′-end ^{32}P-labeled εC/G duplex oligonucleotide. Lane 2, as lane 1, but incubated with 5 μg of *E. coli* crude extracts for 10 min. Lane 3, 3′-end ^{32}P-labeled εC/G duplex oligonucleotide. Lane 4, same as lane 3, but incubated with 5 μg of crude extracts for 10 min. The products of the reaction were analyzed by electrophoresis on a denaturing 20% polyacrylamide gel and visualized by using the PhosphorImager (model Storm 840). For details see *Materials and Methods*.

endonuclease at abasic site. To identify the enzyme responsible for the repair of εC in *E. coli*, we investigated whether this lesion was a substrate for previously characterized DNA repair enzymes. The εCDG activity was checked in crude extracts of various *E. coli* strains deficient in the following DNA repair proteins: RZ1032 (*ung*), GC4803 (*tag*, *alkA*), BH20 (*fpg*), M182 (*micA*), BW372 (*nth*), BW528 (*nfo*, *xth*), BL101 (*uvrA*), NR9288 (*mutD5*), DH5α (*recA*), and JC7623 (*recBC sbcB sbcC*). All of the *E. coli* strains tested contained an activity incising the εC-34/G duplex oligonucleotide (data not shown), suggesting that the εCDG was an uncharacterized enzyme. Hence, the εC repair activity was further purified from *E. coli* cell extract, to identify the protein catalyzing this activity.

Purification of εC-Repair Activity from *E. coli* Cell Extract. The εC repair activity excising εC from the εC-34/G was purified from *E. coli* strain RZ1032 (*ung*, *dut*), using the purification procedure described in *Materials and Methods*, including various columns that allowed us to separate the proteins according to their charge and hydrophobic properties. In the final step, the εC repair activity eluted from the Mono S column as a single symmetrical peak at a concentration of 0.43 M NaCl (Fig. 2*A*). Fractions containing the εC repair activity were analyzed by SDS/PAGE. The most active fraction was found to be purified to apparent homogeneity and consisted of a single polypeptide with an apparent size of about 22 ± 2 kDa (data not shown). This protein was used to obtain peptides for microsequencing.

εCDG Is an Enzymatic Activity of the *E. coli* dsUDG. The partial trypsin proteolysis of εCDG yielded several peptides that were separated by HPLC. Two peptides were subjected to microsequencing. Their N-terminal residues were **QLKPQEAHLLDYR** and **VIYQAGFTDR**. Data bank searches of *E. coli* ORFs revealed that those peptide sequences correspond to the amino acid sequence of a 168-aa protein with a molecular weight of 18,673 Da (ref. 36 and G. Plunkett 3rd, deposited in GenBank in 1995, *ECU*28379, accession no. U28379) that is

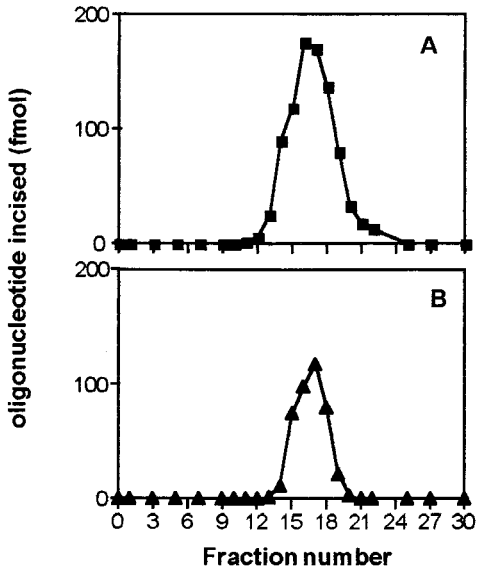

FIG. 2. Distribution of εC and uracil-DNA glycosylase activities after FPLC chromatography on Mono S FPLC HR 5/5 column. The proteins containing the εCDG activity (fraction IV) were loaded on a Mono S FPLC HR 5/5 column and eluted by a linear NaCl gradient. (*A*) 250 fmol of the εC-34/G (■) duplex oligonucleotides were incubated with 1 μl from each column fraction for 10 min in a 50-μl reaction mixture. (*B*) 250 fmol of the U-34/G (▲) duplex oligonucleotides were incubated with 1 μl from each column fraction for 30 min in a 50-μl reaction mixture. The products of the reaction was analyzed as described in Fig. 1 and quantified by using the PhosphorImager (model Storm 840). For details see *Materials and Methods*.

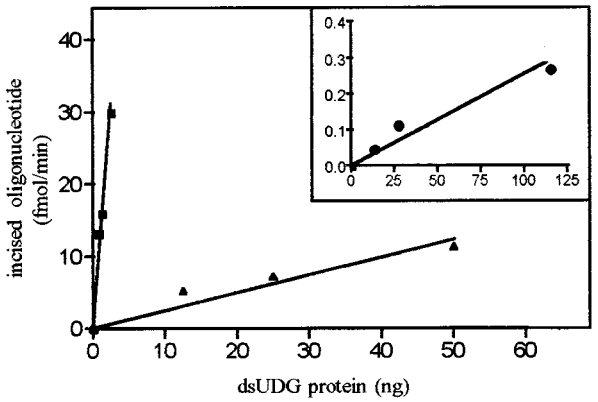

FIG. 3. Activity of the *E. coli* dsUDG protein using as substrate εC-34/G (■), U-34/G (▲,) or T-19/G (●, *Inset*) as substrate. 5′-^{32}P-labeled duplex oligonucleotide containing εC-34/G (■), U-34/G (▲), or T-19/G (●) mismatches (1 pmol in 100-μl reaction mixture) was incubated with increasing amounts of pure dsUDG protein for different periods of time in the presence of the Fpg protein. The products of the reaction were analyzed and quantified as described in Fig. 2. Each point represents the initial velocity of the enzymatic reaction. For details see *Materials and Methods*. Note the difference of scales in the *Inset*.

the dsUDG (24). This enzyme excises uracil residues from duplex DNA containing U/G mispair. Therefore, we have tested whether the εCDG and the dsUDG activities cochromatographed during the purification procedure. As shown in Fig. 2*B*, for the last step of the purification procedure, it is indeed the case.

The kinetics of excision of εC and uracil residues present in a double-stranded oligonucleotide were compared. The results presented in Fig. 3 show that εC residues are excised 54-fold faster than uracil residues in the same experimental conditions. It should be recalled that the preparation of *E. coli* εCDG contains only the dsUDG activity, because it was purified from an *E. coli ung*$^-$ strain. From these results, we conclude that εCDG and dsUDG are two different enzymatic activities of the 168-aa protein, coded for in *E. coli* by the genetic locus *ECU*28379 (G. Plunkett 3rd, deposited in GenBank in 1995, accession no. U28379), and that the εCDG is a catalytic activity of the dsUDG protein that was yet unidentified.

hTDG Excises εC When Present in DNA. Because there is a high extent of homology between amino acid sequences of the dsUDG and the hTDG, we investigated whether the εCDG detected in human cell extracts (22) could be attributed to hTDG. Hence, the hTDG was purified from *E. coli* overexpressing the cloned hTDG by using as substrate an oligonucleotide containing the G/T mismatch. Because the substrate specificity's of the hTDG and its bacterial homologue, the dsUDG protein, overlap, we have verified that the hTDG preparation was not contaminated by the dsUDG protein. The hTDG and dsUDG proteins can be completely separated during the purification procedure. During the first step of purification (the column of DEAE-Sepharose Fast Flow Resin), under the conditions used, the dsUDG protein sticks to the resin, whereas the hTDG protein is recovered in the flow-through fraction. Moreover we have established that during the next step (POROS 20 HS column, a cation exchange resin), any remaining dsUDG activity would have been efficiently separated from the hTDG activity by the linear gradient of NaCl, the dsUDG and the hTDG eluting at 0.35 and 0.50 M, respectively (data not shown). The human enzyme preparation could efficiently excise εC residues from the εC-34/G duplex oligonucleotide, when this residue is situated opposite to G (compare lanes 1 and 3 in Fig. 4). Because there are several examples showing that a modified base is recognized by more than one repair protein (37, 38) and that the efficacy of the

enzymatic activity of a protein for a given substrate depends on its origin (20), we have investigated 11 pure repair proteins from *E. coli* and mammalian origin, involved in base excision repair, as potential candidates to repair εC residues. As shown in Fig. 4, only incubation of the εC-34/G with either the dsUDG or the hTDG, followed by Fpg treatment, leads to the incision of the duplex εC-34/G at the position of the modified base. AlkA, ANPG40, Fpg, Nth, Nfo, Xth, UNG, APDG60, and Tag proteins, although used in massive amounts to detect even marginal activity, do not act on the εC-34/G duplex under our experimental conditions (Fig. 4).

***E. coli* dsUDG and hTDG Proteins Act on DNA Containing εC Residues as DNA Glycosylases.** As shown in Fig. 5, treatment by enzymes nicking at the apurinic (AP) site (Fpg, Nth, or Nfo proteins) does not incise the εC-34/G duplex oligonucleotide (lanes 2–4), showing that there is no detectable AP site in the substrate. Upon treatment with dsUDG (lane 5) or hTDG (lane 9) proteins, no cleavage is detected, showing that our preparations are free of contaminant enzymes nicking at AP sites. However, treatment with enzymes that incise DNA at AP sites, after the action of dsUDG or hTDG proteins, generates a band migrating at the position of the 19-mer (lanes 6–8 and 10–12). Because the various enzymes used to nick at the AP site act through different catalytic mechanisms (39), the products of the enzymatic incision at the AP site of the duplex oligonucleotide migrate differently. The Fpg enzyme treatment cleaves the oligonucleotide at the AP site by a β-δ elimination mechanism generating a ^{32}P-labeled fragment carrying a phosphate at the 3′-end (lanes 6 and 10). Nfo protein, which incises on the 5′-side of the AP site by a hydrolytic mechanism, generates a fragment that has a 3′OH termini and thus migrates slower than the fragment carrying a phosphate at the 3′-end (lanes 8 and 12) (40). The Nth protein incises at the abasic site by a β-elimination mechanism leaving on the 3′ end of the labeled 19-mer an α,β-unsaturated aldehyde that migrates slightly slower than the product of hydrolysis by Nfo (lanes 7 and 11) (40). These experiments show that the dsUDG and hTDG proteins generate AP sites in DNA containing εC residues and therefore act as DNA glycosylases.

In addition, the nature of the product excised from εC-4/G by the repair enzymes was characterized by analysis on HPLC as described in *Materials and Methods*. The product of excision eluted as a single peak at 11.9 min. Furthermore this product cochromatographed with an authentic sample of ethenocytosine. No material was detected at the position of the nucleoside (data not shown). These two independent experiments show that εC residues are excised by a DNA glycosylase mechanism.

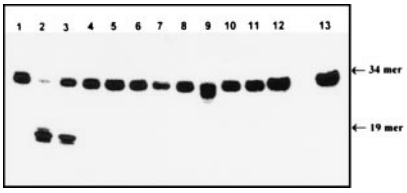

FIG. 4. Action of various *E. coli* and human DNA repair proteins on the 34-mer duplex εC/G oligonucleotide. The 5′-^{32}P-labeled εC-34/G was incubated with an excess of the various pure repair protein at 37°C for 30 min (unless otherwise stated). Except for the control εC-34/G oligonucleotide and this oligonucleotide treated with Nth, Nfo, or Xth protein, the reactions were made in the presence of Fpg protein (50 ng) to reveal any abasic site generated by DNA glycosylases devoid of β-lyase activity. Lane 1, control εC-34/G oligonucleotide. Lane 2, as lane 1, but treated by *E. coli* dsUDG protein (5 ng). Lane 3, hTDG (150 ng, 30°C). Lane 4, AlkA (400 ng). Lane 5, ANPG40 (1.3 μg). Lane 6, Fpg protein (1 μg). Lane, 7, Nth protein (100 ng). Lane 8, Nfo (1.2 μg). Lane 9, Xth (4 nM, 10 min, 23°C). Lane 10, UNG (85 ng). Lane 11, APDG60 protein (1 μg). Lane 12, Tag I (350 ng). Lane 13, control as 1. The products of the reaction were analyzed as described in Fig. 1. For details see *Materials and Methods*.

The *E. coli* dsUDG Processes G/T Mismatch. The ability of the hTDG and dsUDG to excise εC suggests that both proteins could have a very similar substrate specificity. Because hTDG recognizes substrates containing G/T, U/G, and εC/G mismatches, we investigated whether the dsUDG protein is endowed with a G/T mismatch-specific activity. To measure this activity, we used, as suggested by Sibghat-Ullah *et al.* (41), an oligonucleotide T-19/G where the T residue is in the context of TpG/T. In fact, (see Fig. 3 *Inset*) the dsUDG processes G/T mispairs in DNA, although much less efficiently than the U- and εC-containing substrates.

Comparison of the Kinetic Parameters for Excision of εC, Uracil, and Thymine from Duplex Oligonucleotides Containing εC/G, U/G, and G/T Mispairs, Respectively, by the dsUDG and the hTDG. To evaluate the relative substrate specificity of the dsUDG for its various substrates, we measured the kinetic parameters for the excision of εC, uracil, and thymine residues, opposite to G in a double-stranded oligonucleotide. The results show that the best substrate for the *E. coli* dsUDG enzyme is, by far, εC as compared with uracil, because the specificity constant is 50 times lower for the latter (Table 1).

In addition, Table 1 presents the kinetic constants measured for the hTDG acting on the same set of oligonucleotides. The comparison of the specificity constants for the enzyme using as substrate εC/G or U/G mismatches leads to the conclusion that for the human enzyme uracil and εC are equally good substrates and the most preferred substrates as compared with T/G mismatch.

Base Pair Specificity of the Ethenocytosine-DNA Glycosylases. The dsUDG and the hTDG recognized εC only when present in a double-stranded oligonucleotide. No detectable excision of εC was observed when the lesion was present in a single-stranded oligonucleotide (data not shown). The base-pair specificity of both enzymes was measured by using duplex

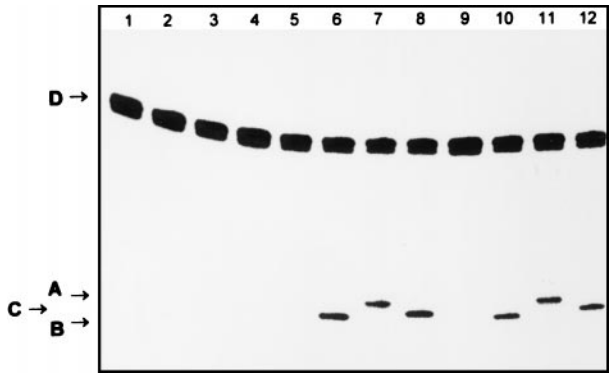

FIG. 5. Mechanism of action of the dsUDG and hTDG proteins on εC/G oligonucleotide. The εC-34/G duplex oligonucleotide was incubated with dsUDG protein or hTDG protein, and subsequently treated or not with proteins nicking at AP sites to reveal abasic sites generated by dsUDG or hTDG proteins. The 5′-^{32}P-labeled 34-mer εC-34/G duplex oligonucleotide is: lane 1, incubated at 37°C for 30 min; lanes 2, 6, and 10, incubated with 100 ng of Fpg protein at 37°C for 10 min; lanes 3, 7, and 11, incubated with 100 ng of Nth protein at 37°C for 10 min; lanes 4, 8, and 12, incubated with 100 ng of Nfo protein at 37°C for 10 min; lanes 5–8, incubated with 2 ng of dsUDG protein at 37°C for 10 min; lanes 9–12, incubated with 50 ng of hTDG protein at 30°C for 30 min. The products of the reaction were analyzed as described in Fig. 1. Arrow A indicates the 19-mer oligonucleotide containing an α,β-unsaturated aldehyde at the 3′-end. Arrow B indicates the 19-mer oligonucleotide containing a phosphate at the 3′-end. Arrow C indicates the 19-mer oligonucleotide containing 3′-OH termini. Arrow D indicates the 34-mer εC-34 oligonucleotide. For details see *Materials and Methods*.

Table 1. Kinetic constants of the *E. coli* dsUDG and hTDG proteins for the excision of εC, uracil, and thymine (opposite to guanine)

Enzyme	Substrate*	K_m, nM	k_{cat}, min^{-1}	k_{cat}/K_m, min × nM^{-1}
dsUDG	εC/G†	2.5 ± 1.6	0.95 ± 0.22	0.38
	U/G	22.7 ± 8.6	0.17 ± 0.03	0.77×10^{-2}
	T/G	26 ± 10	$(0.43 \pm 0.06) \times 10^{-3}$	0.17×10^{-4}
hTDG	εC/G	24.3 ± 14.9	$(9.2 \pm 1.7) \times 10^{-3}$	0.38×10^{-3}
	U/G	12 ± 5	$(21 \pm 3) \times 10^{-3}$	1.73×10^{-3}
	T/G	12.8 ± 4.9	$(0.9 \pm 0.1) \times 10^{-3}$	0.071×10^{-3}

*Substrate concentration range 2.5–100 nM.
†Substrate concentration range 0.1–6 nM.

oligonucleotides containing mismatches generated by each of the four different bases opposite εC or uracil. In each case, the initial velocity of the excision of εC or uracil was measured. The results are presented in Table 2.

The excision of εC by the dsUDG or by hTDG does not show any strict preference, although the εC/G mismatch was the best substrate. At variance, the repair of uracil residues exhibits a marked preference according to the opposite base. In the case of the *E. coli* protein, uracil is excised from U/G, U/T, and U/C mismatches with a similar efficiency, whereas its excision from U/A is negligible. The hTDG protein strongly preferred U/G as a substrate, the repair of uracil in U/C is very negligible, and it is not excised from U/T and U/A. The striking fact is that εC residues are excised from all four mismatches with a comparable efficiency by both proteins.

DISCUSSION

Ethenobases have attracted much attention as critical candidates in the etiology of human cancers, because these adducts lead to misincorporation upon replication or transcription. Ethenoadducts in DNA could be formed either by exogenous sources (vinyl chloride, ethyl carbamate, etc.) or by products of lipid peroxidation such as *trans*-4-hydroxy-2-nonenal generated during the cellular metabolism (9, 10).

Because ethenobases are known to be promutagenic and genotoxic (12, 13, 15–17), they have to be removed from the genomic DNA. In *E. coli* and in *Saccharomyces cerevisiae*, the enzymatic activity excising εA has been identified as the 3-methyladenine-DNA glycosylase, the AlkA and Mag proteins, respectively (20), whereas proteins excising εC from these two organisms are unknown. Enzymatic activities excising εC and εA have been identified in mammalian crude cell extracts (18). The human protein binding to and excising εA has been partially purified (19) and identified as the ANPG protein (19, 20). Using partially purified proteins from HeLa cells and ANPG knockout mice, it has been shown that the enzymes excising εA and εC are two different proteins (22, 23).

In the present study, an εCDG has been purified to homogeneity from *E. coli* cells extract and was identified as the *E. coli* dsUDG (24), which is the homologue of the hTDG (27). However, the molecular weight of the purified εCDG/dsUDG, estimated by its mobility on PAGE under denaturing conditions (22 ± 2 kDa), differs from the calculated one based on the gene sequence (18.7 kDa). Moreover we have established that hTDG, the human homologue of dsUDG, also excises εC residues.

To compare the substrate specificity of the *E. coli* and the human enzymes, the kinetic parameters were measured for both enzymes, using three different duplex oligonucleotides containing different mismatches. The dsUDG protein acted on DNA substrates in the order εC/G ≫ U/G ≫ G/T. It should be emphasized that the most preferred substrate for dsUDG is, by far, εC. The enzyme has an extremely efficient kinetic constant acting on εC/G (k_{cat}/K_m = 0.4 min·nM^{-1}), 52-fold higher than on U/G mismatch ($k_{cat}/K_m = 0.77 \times 10^{-2}$ min·nM^{-1}), which is believed to be the physiological substrate of this enzyme. This suggests, if one extrapolates the *in vitro* results to the *in vivo* activity, that the main role of that glycosylase *in vivo* could be the repair of εC lesions in DNA. The physiological relevance of the U/G activity of the dsUDG has not yet been ascertained.

Moreover, the human protein hTDG, previously characterized as an enzyme repairing T in G/T mismatches, very efficiently repairs εC residues in DNA. The order of the hTDG preference for different DNA substrates is U/G > εC/G > T/G. The difference among the kinetics parameters measured for the excision of εC residues is less prominent for the human enzyme as compared with the dsUDG. This situation is different from what has been observed for the excision of εA, the mammalian enzymes (human and rat) being much more efficient than the prokaryotic ones (20). From the comparison of the k_{cat}/K_m values of the hTDG protein for each substrate, we conclude that uracil and εC are the most preferred substrates. The specificity of the human protein for uracil mismatch is 4.5 times higher than for εC, the k_{cat}/K_m value for thymine being much lower than for uracil and εC.

The lower k_{cat}/K_m value measured for the hTDG using an εC substrate, compared with the constant measured for the dsUDG, suggests that *in vivo* this enzyme may act as part of an efficient multiprotein complex. Ascertaining the role of hTDG *in vivo* will require further investigations using genetic approaches. The high promutagenic properties of εC in mammalian cells (13) and its repair by hTDG lead to the possibility of identifying a human genetic disease associated with a deficient hTDG activity.

The excision of T in a G/T mismatch by dsUDG adds a substrate to this enzyme. The extremely low constant measured ($k_{cat}/K_m = 1.7 \times 10^{-5}$) suggests that this enzymatic activity does not have a real biological significance. However, this substrate, in conjunction with the known structure of

Table 2. The influence of the base opposite to εC and/or uracil residues on the rates of excision by *E. coli* dsUDG and hTDG proteins

Substrate	DNA glycosylase activity, %	
	E. coli dsUDG	hTDG
εC/G	100	100
εC/A	46	26.5
εC/T	93	21
εC/C	61	71
U/G	100	100
U/A	1	<0
U/T	84	<0
U/C	94	2

The εC-34 mer and/or U-34 mer were annealed with the complementary 34-mer oligonucleotides to generate the following mismatches: εC/G, εC/A, εC/T, and εC/C and/or U/G, U/A, U/T, U/C. Enzymatic activity was measured as a function of time, and the initial velocities were determined. For details see *Materials and Methods*.

duplexes containing εC/T, εC/G, or εC/A mismatches (42–44), will be of interest for structural investigations aiming to elucidate the molecular mechanisms involved in the catalytic action of the dsUDG. Therefore, the human and the *E. coli* enzymes exhibit the same substrate specificity.

Repair of thymine and uracil residues by dsUDG and hTDG proteins strongly depends on the nature of the opposite base (see Table 2) (45). At variance, the repair of the εC residue is not markedly influenced by the opposite base. It should be noted that the efficiency of the excision of the εA residues by AlkA, MAG, and ANPG proteins does not depend on the nature of the opposite base (20). These facts raise the possibility that the processes involved in the recognition of the various modified bases could be different, the εC residue being recognized *per se* whereas the uracil and thymine residues are recognized through the structure of the mismatch. In fact, the U/A and the T/A are not repaired (Table 2 and ref. 45). One can speculate that the dsUDG and the hTDG directly recognize the ethenoadduct and flip out the εC residue into a specific pocket within their active site, similar to mechanism described for the UNG protein (46).

In conclusion, our results show that the DNA repair activity excising εC is associated in *E. coli* with the dsUDG and in human cells with the thymine-DNA glycosylase. The kinetics data demonstrate that εC is the most preferred substrate for both enzymes. The association of the mismatch-specific DNA glycosylase activity with the εCDG activity in the same protein is conserved during evolution, because it is observed in two unrelated species, human and *E. coli*. This result implies that a possible role of the human and bacterial enzymes *in vivo* could be to protect the integrity of the cellular genome from carcinogens and mutagens producing εC adducts in DNA.

We thank Dr. J. Jiriçny for the gift of the plasmid expressing the human pT7-hTDG gene and Dr. J. Derancourt (UPR 9008 Centre National de la Recherche Scientifique, Montpellier) for the determination of the amino acid sequence of the εCDG/dsUDG. We also thank Dr. O. Fedorova for invaluable help. This work was supported by grants from the European Commission (ENV4-CT97-0505), Comité Radioprotection-Electrecité de France, and Fondation Franco-Norvegienne. M.S. is the recipient of a grant from the Fondation pour la Recherche Médicale.

1. Green, T. & Hathway, D. E. (1978) *Chem. Biol. Interact.* **22,** 211–224.
2. Miller, J. A. & Miller, E. C. (1983) *Br. J. Cancer* **48,** 1–15.
3. Guengerich, F. P. & Kim, D.-H. (1991) *Chem. Res. Toxicol.* **4,** 413–421.
4. Dahl, G. A., Miller, J. A. & Miller E. C. (1978) *Cancer Res.* **38,** 3793–3804.
5. Leithäuser, M. T., Liem, A., Steward, B. C., Miller, E. C. & Miller J. A. (1990) *Carcinogenesis* **11,** 463–473.
6. Park, K. K., Surh, Y. J., Steward, B. C. & Miller, J. A. (1990) *Biochem. Biophys. Res. Commun.* **169,** 1094–1098.
7. Bolt, H. M. (1988) *Crit. Rev. Toxicol.* **18,** 299–309.
8. Guengerich, F. P., Min, K. S., Persmark, M., Kim, M. S., Humphreys, W. G. & Cmarik, J. L. (1994) in *DNA Adducts: Identification and Significance*, eds. Hemminki, K., Dipple, A., Shuker, D. E. G., Kadlubar, F. F., Segerbäck, D. & Bartsch, H. (International Agency for Research on Cancer, Lyon, France), pp. 57–72.
9. Chung, F.-L., Chen, H.-J. C. & Nath, R. G. (1996) *Carcinogenesis* **17,** 2105–2111.
10. El-Ghissassi, F., Barbin, A., Guichard, Y. & Bartsch, H. (1995) *Chem. Res. Toxicol.* **8,** 278–283.
11. Marnett, L. J. & Burcham, P. C. (1993) *Chem. Res. Toxicol.* **6,** 785–771.
12. Basu, A. K., Wood, M. L., Niedernhofer, L. J., Ramos, L. A. & Essigmann, J. M. (1993) *Biochemistry* **32,** 12793–12801.
13. Moriya, M., Zhang, W., Johnson, F. & Grollman, A. P. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 11899–11903.
14. Wood, M. L., Esteve, A., Morningstar, M. L., Kuziemko, G. M. & Essigmann, J. M. (1992) *Nucleic Acids Res.* **20,** 6023–6032.
15. Pandya, G. A. & Moriya, M. (1996) *Biochemistry* **35,** 11487–11492.
16. Cheng, K. C., Preston, B. D., Cahill, D. S., Dosanjh, M. K., Singer, B. & Loeb, L. A. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 9974–9978.
17. Langouët, S., Müller, M. & Guengerich, F. P. (1997) *Biochemistry* **36,** 6069–6079.
18. Oesch, F., Adler, S., Rettelbach, R. & Doerjer, G. (1986) in *The Role of Cyclic and Nucleic Acid Adducts in Carcinogenesis and Mutagenesis*, eds. Singer, B. & Bartsch, H. (Oxford Univ. Press, New York), pp. 373–379.
19. Singer, B., Antoccia, A., Basu, A. K., Dosanjh, M. K., Fraenkel-Conrat, H., Gallagher, P. E., Kusmierek, J. T., Qiu, Z. H. & Rydberg, B. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 9386–9390.
20. Saparbaev, M., Kleibl, K. & Laval, J. (1995) *Nucleic Acids Res.* **23,** 3750–3755.
21. Matijasevic, Z., Sekiguchi, M. & Ludlum, D. B. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 9331–9334.
22. Hang, B., Chenna, A., Rao, S. & Singer, B. (1996) *Carcinogenesis* **17,** 155–157.
23. Hang, B., Singer, B., Margison, G. P. & Elder, R. H. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 12869–12874.
24. Gallinari, P. & Jiriçny, J (1996) *Nature (London)* **383,** 735–738.
25. Seeberg, E., Eide, L. & Bjoeras, M. (1995) *Trends Biochem. Sci.* **20,** 391–397.
26. Slupska, M. M., Baikalov, C., Luther, W. M., Chiang, J.-H., Wei, Y.-F. & Miller, J. H. (1996) *J. Bacteriol.* **178,** 3885–3892.
27. Neddermann, P., Gallinari, P., Lettieri, T., Schmid, D., Truong, O., Hsuan, J. J., Wiebauer, K. & Jiriçny, J. (1996) *J. Biol. Chem.* **271,** 12767–12774.
28. Saparbaev, M. & Laval, J. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 5873–5877.
29. Boiteux, S., O'Connor, T. R., Lederer, F., Gouyette, A. & Laval, J. (1990) *J. Biol. Chem.* **265,** 3916–3922.
30. Levin, J. D., Johnson, A. W. & Demple, B. (1988) *J. Biol. Chem.* **263,** 8066–8071.
31. Graves, R. J., Felszenszwalb, I., Laval, J. & O'Connor, T. R. (1992) *J. Biol. Chem.* **267,** 14429–14435.
32. O'Connor, T. R. & Laval, F. (1990) *EMBO J.* **9,** 3337–3342.
33. Tudek, B., VanZeeland, A. A., Kusmierek, J. T. & Laval, J. (1998) *Mutation Res.*, in press.
34. O'Connor, T. & Laval, J. (1991) *Biochem. Biophys. Res. Commun.* **176,** 1170–1177.
35. O'Connor, T. R. (1994) *Nucleic Acids Res.* **21,** 5561–5569.
36. Borodovsky, M., McIninch, J. D., Koonin, E. V., Rudd, K. E., Medigue, C. & Danchin, A. (1995) *Nucleic Acids Res.* **23,** 3554–3562.
37. Hatahet, Z., Kow, Y. W., Purmal, A. A., Cunningham, R. P. & Wallace, S. (1994) *J. Biol. Chem.* **269,** 18814–18820.
38. Melamede, R. J., Hatahet, Z., Kow, Y. W., Ide, H. & Wallace. S. S. (1994) *Biochemistry* **33,** 1255–1264.
39. Dodson, M. M., Michaels, M. L. & Lloyd, R. S. (1994) *J. Biol. Chem.* **269,** 32709–32712.
40. Bailly, V., Verly, W. G., O'Connor, T. & Laval, J. (1989) *Biochem. J.* **262,** 581–589.
41. Sibghat-Ullah, Gallinari, P., Xu, Y.-Z., Goodman, M. F., Bloom, L. B., Jiriçny, J. & Day, R. S., III (1996) *Biochemistry* **35,** 12926–12932.
42. Korobka, A., Cullinan, D., Cosman, M., Grollman, P. A., Patel, D. J., Eisenberg, M. & de los Santos, C. (1996) *Biochemistry* **35,** 13310–13318.
43. Cullinan, D., Korobka, A., Grollman, P. A., Patel, D. J., Eisenberg, M. & de los Santos, C. (1996) *Biochemistry* **35,** 13319–13327.
44. Cullinan, D., Johnson, F., Grollman, P. A., Eisenberg, M. & de los Santos, C. (1997) *Biochemistry* **36,** 11933–11943.
45. Neddermann, P. & Jiriçny, J. (1993) *J. Biol. Chem.* **268,** 21218–21224.
46. Savva, R., McAuley-Hecht, K., Brown, T. & Pearl, L. (1995) *Nature (London)* **373,** 487–493.