A 55-kDa protein isolated from human cells shows DNA glycosylase activity toward 3,N^4-ethenocytosine and the G/T mismatch

B. HANG, M. MEDINA, H. FRAENKEL-CONRAT, AND B. SINGER*

Donner Laboratory, Lawrence Berkeley National Laboratory, University of California, Berkeley, CA 94720

Contributed by H. Fraenkel-Conrat, September 14, 1998

ABSTRACT Etheno adducts in DNA arise from multiple endogenous and exogenous sources. Of these adducts we have reported that, 1,N^6-ethenoadenine (eA) and 3,N^4-ethenocytosine (eC) are removed from DNA by two separate DNA glycosylases. We later confirmed these results by using a gene knockout mouse lacking alkylpurine-DNA-N-glycosylase, which excises eA. The present work is directed toward identifying and purifying the human DNA glycosylase activity releasing eC. HeLa cells were subjected to multiple steps of column chromatography, including two eC-DNA affinity columns, which resulted in >1,000-fold purification. Isolation and renaturation of the protein from SDS/polyacrylamide gel showed that the eC activity resides in a 55-kDa polypeptide. This apparent molecular mass is approximately the same as that reported for the human G/T mismatch thymine-DNA glycosylase. This latter activity copurified to the final column step and was present in the isolated protein band having εC-DNA glycosylase activity. In addition, oligonucleotides containing εC-G or G/εT, could compete for εC protein binding, further indicating that the εC-DNA glycosylase is specific for both types of substrates in recognition. The same substrate specificity for εC also was observed in a recombinant G/T mismatch DNA glycosylase from the thermophilic bacterium, Methanobacterium thermoautotrophicum THF.

The four etheno adducts of DNA and RNA have been of considerable interest to organic chemists and physical scientists due to their physical, chemical and spectroscopic properties which had broad applications in protein-nucleic acid interactions and DNA structure [reviewed by Leonard (1, 2) and refs therein]. These adducts became of major interest when they were found to be formed by a variety of environmental agents, as well as produced endogenously (3–7). Mutagenesis studies have shown a wide range of mutagenic frequency depending on the type of the adduct, type of mutation and the system used for detection and quantitation (8–19).

For more than one decade, this laboratory has focused on studies on the repair (20–26) and replication/transcription (8–12) of etheno derivatives of dA, dC, and dG. The differing structures of the etheno adducts and their effect on base pairing and base stacking (27–31) influences both repair and replication. Much of the data has been obtained by using prokaryotic systems, which are not always identical to those now found in the more widely used mammalian systems. Although no model experiment can reproduce exactly what occurs in human cells, there is considerable progress on understanding repair of mutagenic lesions by using human cells and tissues (reviewed in ref. 32 and refs. therein). It was established earlier in repair studies, by using a human system, that all four etheno adducts were released by HeLa cell-free extracts, indicating that they are substrates for DNA glycosylases (24). After partial purification from HeLa cells, 3,N^4-ethenocytosine (eC) repair activity was found to be separate from 1,N^6-ethenoadenine (eA) repair activity (25), which is a function of alkylpurine-DNA-N-glycosylase (APNG) (22). A knockout mouse lacking APNG was then used as a genetic approach to verify these in vitro data (26). Under these conditions, of the two etheno adducts tested, only eC was released by the cell-free extracts of such mice, indicating that there was a different gene product for εC repair (26). We also found that the glycosylase responsible for εC recognition had an unusually high molecular mass for a DNA glycosylase, as judged by size exclusion chromatography (32).

The calculated value was close to that reported by Jiricny and coworkers (33, 34) for the human G/T mismatch thymine-DNA glycosylase.

In this work, we report that further purification of the εC glycosylase activity from HeLa cells showed that the εC activity resided in a 55-kDa polypeptide. This protein also had a coexisting G/T mismatch activity. Other glycosylase substrates tested were not cleaved by this purified human εC-DNA glycosylase. Our conclusion agrees with that of Saparbaev and Laval (35), who, after our experiments were completed, reported a similar finding by using a purified recombinant human G/T mismatch thymine-DNA glycosylase.

MATERIALS AND METHODS

Oligonucleotide Substrates. The sequences of oligonucleotides used in this study are listed in Fig. 1. The synthesis of the 3,N^4-ethenodeoxyctydine phosphoramidite and its incorporation into oligomers has been described elsewhere (36, 24). The sequence of the 45-mer duplex containing G/T (sequence 4) was previously described by Sibghat-Ullah et al. (37). The oligomers were synthesized by using an Applied Biosystem model 392 DNA synthesizer and purified by HPLC and/or PAGE.

Repair Enzymes. Recombinant thermostable G/T mismatch glycosylase was purchased from Trevigen (Gaithersburg, MD). Uracil-DNA glycosylase was obtained from GIBCO/BRL.

Preparation of εC-DNA Affinity Matrix. The εC-containing 16-mer (Fig. 1, sequence 2, upper strand) was annealed to a complementary strand to produce sticky ends. Then 260 mg of the duplex was lightly labeled with 5 μCi [γ-32P]ATP (specific activity 6,000 Ci/mmol; 1 Ci = 37 GBq; Amersham). The ligation was carried out in the buffer containing 50 mM Tris-HCl (pH 7.8), 10 mM DTT, 10 mM MgCl2, 1 mM

Abbreviations: εA, 1,N^6-ethenoadenine; εC, 3,N^4-ethenocytosine; m^3A, 3-methyladenine; U, uracil; AP, apurinic/apyrimidinic; APNG, alkylpurine-DNA-N-glycosylase; FPLC, fast protein liquid chromatography.

*To whom reprint requests should be addressed.
For detection of eC activity:
1. 5'-CGCCATATGCGTTACGCGGCTGGAAT
   GGGATGC CCCCAGTGGCTGAGCT-5'

For eC-DNA affinity matrix:
2. 5'-GAAGCTATCGCGGGTAC
CGATC GCCCATGCTAG-5'

For detection of G/T(U) mismatch activity:
3. 5'-CGCCATATGCTGTAACCGGCTGGAAT
   GGGATGC CCCCAGTGGCTGAGCT-5'

4. 5'-ATTAGGATACGACACCCGCTGGGGCACCACCC
   GATGAAGGGCTGGCTGAGCCTGAT-5'

5. 5'-CGCCATAGUGGGTACCGGCTGGAAT
   GGGATGC CCCCAGTGGCTGAGCT-5'

Unmodified oligomer duplex for competition:
6. 5'-CGCCATATGCGGTACCGGCTGGAAT
   GGGATGC CCCCAGTGGCTGAGCT-5'

7. 5'-GAAGCTATCGCGGGTACGCGGCTGGAAT
   CATGAGCAAGGATGC CCCCAGTGGCTGAGCTGAT-5'

Fig. 1. Oligodeoxynucleotides used in this study. The numbers are referred to in the text. Mismatches are indicated in bold type.

hexammine cobalt chloride, 0.25 mM spermidine, 1 mM ATP, 25 μg/ml BSA, and 30 Weiss units of T4 DNA ligase (6,000 units/ml, New England Biolabs) for 7.5 hr at 15°C. The extent of ligation was checked by running PAGE after a 4-h incubation, at which point >90% of the 16-mer duplex was elongated to heterogeneous multimers.

Coupling the eC-DNA to the cyanogen bromide (CNBr)-activated Sepharose CL-4B (Pharmacia) was performed essentially as described (38). The eC-DNA-resin was equilibrated in PC buffer (25 mM Hepes-potassium hydroxide, pH 7.8-8.0, 5 mM EDTA, 0.1 M potassium chloride) with PC buffer, and loaded onto a 1-ml Mono-S HR5/5 FPLC column (Pharmacia) equilibrated in PC buffer. The column was then applied to a Blue Sepharose CL-6B column (8 x 2.5 cm) (Pharmacia), which was packed and pretreated as directed by Rydberg et al. (20, 21).

The active fractions from the phosphocellulose columns were pooled (Table I, fraction III) and desalted by the same filter device described above to lower the conductivity to be equivalent to a KCl concentration of < 200 mM. The sample was then applied to a Blue Sepharose CL-6B column (8 x 2.5 cm) (Pharmacia), which was packed and equilibrated with three volumes of the PC buffer. The column was rinsed with 30 ml of PC buffer and eluted with a linear gradient (100 ml) of 0.2–1 M KCl in PC buffer.

After the Blue Sepharose chromatography, the active fractions were pooled (Table I, fraction IV), concentrated, diluted with PC buffer, and loaded onto a 1-ml Mono-S HR5/5 FPLC column (Pharmacia) equilibrated in PC buffer. The column was washed with 10 ml of PC buffer and eluted with a gradient of 30 ml from 0.1–0.5 M KCl in the buffer followed by a 3-ml gradient from 0.5–1 M KCl at a flowrate of 0.4 ml/min. Seven milliliters of active fractions was collected (Table I, fraction V).

Fraction V was concentrated to 4 ml and diluted 1:3.5 with PC buffer. Then 30 μg of poly(dI-dC)-(dI-dC) was added to the fraction as a nonselective competitor. After incubation at 4°C for 10 min, the sample was loaded, in aliquots, to the 1.5-ml eC-DNA affinity column (see previous section). Approximately 0.5 ml of the sample was allowed to be absorbed into the matrix each time and to stay for 5 min for binding. The column was then washed with 8 ml of PC buffer and developed with a 12-ml gradient from 0.1–0.8 M KCl in PC buffer at a flowrate of 0.3 ml/min. Fractions of 0.5 ml were collected and stored in siliconized tubes. The second eC-DNA affinity chromatography was carried out by using as competitor a 48-bp unmodified oligomer duplex composed of two repeats of the 16-mer eC-oligomer used for the affinity matrix (Fig. 1, duplex 7). A linear gradient (11 ml) from 0.2–0.8 M KCl in PC buffer was applied to elute a 1-ml eC-DNA affinity column. Protein concentrations were determined by Bradford method (39).

SDS/PAGE and Silver Staining. Discontinuous SDS/PAGE was carried out according to Laemmli (40). The gels were stained by using a silver-staining kit according to the manufacturer's instructions (Pharmacia). SDS protein molecular mass markers were from BioRad (size range: 14.4–94.0 kDa).

Isolation and Renaturation of the eC Glycosylase from SDS/PAGE. Sixty microliters of the active fractions from the first eC-DNA affinity column was electrophoresed on a SDS/PAGE (5% stacking and 10% separating gel). The molecular mass markers and 5 μl of the same active fraction were run also in the side lanes and later silver stained. The unstained lane containing the bulky sample was cut into appropriate slices that were then each incubated in an elution buffer (33) overnight with vigorous shaking at 30°C. The denaturation with 6 M guanidine HCl (Sigma) and subsequent renaturation was carried out essentially as described by Hager and Burgess.

Table 1. Purification of eC-DNA glycosylase from HeLa cells

<table>
<thead>
<tr>
<th>Purification step</th>
<th>Volume, ml</th>
<th>Total protein, mg</th>
<th>Total eC activity, units*</th>
<th>Specific activity, units/mg</th>
<th>Yield, %</th>
<th>Purification fold, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Whole cell extract</td>
<td>533</td>
<td>5,785</td>
<td>50,816</td>
<td>8.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II. Ammonium sulfate precipitate</td>
<td>286</td>
<td>3,639</td>
<td>40,841</td>
<td>11.2</td>
<td>80.4</td>
<td>1.3</td>
</tr>
<tr>
<td>III. Phosphocellulose (P11)</td>
<td>161</td>
<td>444</td>
<td>17,124</td>
<td>38.6</td>
<td>33.7</td>
<td>4.4</td>
</tr>
<tr>
<td>IV. Blue Sepharose</td>
<td>30</td>
<td>63.3</td>
<td>5,398</td>
<td>85.3</td>
<td>10.6</td>
<td>9.7</td>
</tr>
<tr>
<td>V. Mono-S</td>
<td>7</td>
<td>9</td>
<td>2,945</td>
<td>327.2</td>
<td>5.8</td>
<td>37</td>
</tr>
<tr>
<td>VI. eC-DNA affinity (1)</td>
<td>4.5</td>
<td>0.2</td>
<td>1,838</td>
<td>9,190</td>
<td>3.6</td>
<td>1,044</td>
</tr>
<tr>
<td>VII. eC-DNA affinity (2)</td>
<td>3.5</td>
<td>nd†</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*One unit of eC-DNA glycosylase is defined as 1 fmol of the eC-oligomer cleaved after 10 min at 37°C.
†No measurable protein was recovered but there was sufficient eC activity to determine that G/T mismatch activity remained.
by introducing eC-DNA affinity chromatography. Table 1 summarizes the steps used in the purification. The relative activities of eC fractions were monitored by using a nicking assay described in Materials and Methods.

The current purification scheme started with whole cell extracts from 105 g of HeLa cells. The initial stability study using the (NH₄)₂SO₄ fraction showed that eC activity is relatively stable with remaining activity of 95% (−20°C), 81% (4°C), 57% (20°C), and 24% (37°C) after 24 hr in PC buffer (activity at −75°C is treated as 100%). (NH₄)₂SO₄ precipitation and filtration through a cation-exchange phosphocellulose column removed 92% of the protein, and there was a 4.4-fold enrichment of the specific activity (Table 1). The eC protein bound strongly to the cation P11 matrix, and one single peak of activity was observed (Fig. 2A). The use of a dye affinity column, Blue Sepharose CL-6B, yielded a further 10-fold enrichment of the activity. Moreover, the volume of the sample was reduced to 30 ml, which allowed the use of 1-ml Mono-S FPLC column in the next step (Table 1, step V). The distribution profile of eC activity from this column is shown in Fig. 2B. Mono-S FPLC cation-exchange chromatography produced an overall 37-fold increase in purification. This step removed contaminating 5’ AP endonuclease(s) as judged by the cleavage pattern on denaturing PAGE (data not shown).

The next step of purification, eC-DNA affinity (1) chromatography (Table 1, step VI), produced the greatest increase in specific activity, 1,044-fold. We previously demonstrated an eC-binding protein in HeLa crude extracts and partially purified fractions (24, 25). In this work, the binding activity was found to copurify with the eC-nicking activity as shown in Mono-S fractions (Fig. 3A). The protein-DNA complex was shown to be stable for at least 1 hr at 4°C and was damage (eC)-specific (Fig. 4). We then could take advantage of these properties and prepare the eC-DNA affinity matrix. In the first

Fig. 2. Autoradiogram of 12% denaturing PAGE showing eC activity profile from two column steps of purification. (A) Fractionation of eC activity by using phosphocellulose P11 chromatography (Table 1, step III). The cleavage of the 25-mer eC-oligomer (Fig. 1, seq. 1) is shown as indicated by the 7-mer size marker on the left. The peak of the activity elutes at 0.9 M KCl. (B) eC activity profile from Blue Sepharose CL-6B chromatography (Table 1, step IV). The peak of the activity elutes at 0.7 M KCl. In both A and B, 3-ml fractions were collected and 2.5 μl of each was tested in a total of 10 μl of reaction at 37°C for 1 hr as described in Materials and Methods. Ft, flow through.
round of purification, eC nicking activity was detected in fractions eluted from 0.45–0.65 M KCl (Fig. 5B) and was in agreement with the profile of the eC-binding activity in the same fractions (Fig. 5B). The second eC-DNA affinity chromatography (Table 1, step VII) led to a very low recovery of the eC activity so that the protein concentrations in all fractions were undetectable even under the silver staining. However, there was detectable eC activity at the same KCl concentration eluting eC activity in the first affinity column. The apparent molecular mass of the eC activity was determined by isolating the protein bands from SDS-PAGE of the first eC-DNA affinity fractions and subsequent renaturation of the eluted proteins (Materials and Methods). The enzymatic test indicated that the target band is ~55 kDa (Fig. 6, arrow). This size is unusual for DNA glycosylases, which are generally between 25–40 kDa (44). We had earlier reported a similarly large size (~66 kDa) for the same activity by using gel filtration chromatography (32). This apparent discrepancy in size for the same activity is attributed to the methods used.

We previously reported the eC release from DNA by HeLa cell-free extracts (24). In this study, a nicking assay by using 3’-end-labeled substrate oligomers and purified fractions was performed to confirm this glycosylase-mediated mechanism. Indeed, the same cleavage pattern was found when the eC oligomer was treated with purified fractions, as when a uracil-containing oligomer was treated with uracil-DNA glycosylase, suggesting that the same mechanism was used by eC activity as that for uracil-DNA glycosylase (data not shown).

**Coexistence of eC and G/T Mismatch Repair Activity.** As stated above, the molecular mass for eC activity was close to the unusually large size of the purified and cloned G/T mismatch thymine-DNA glycosylase (33, 34). We then tested the newly purified eC-DNA glycosylase for possible G/T mismatch activity in the course of purification. The coelution
of these two activities was found in the last three column runs. The superimposed peak activities for steps V and VI are shown in Fig. 5.

The isolated 55-kDa polypeptide in Fig. 6 also showed activity toward both εC and G/T mismatch substrates. To test the possibility that two polypeptides with same size coeluted, competition assays were performed in which the εC binding of first affinity fraction was competed with oligomers containing εC-G, G/T, or G/U. All the mismatches were specific competitors with εC-G and G/U the most efficient (data not shown). G/U, in double stranded form, is also a substrate for the human G/T mismatch glycosylase and showed better protein binding than G/T (33). We thus concluded that a single repair protein had both activities. Finally, a commercial β-Lactamase (EC 3.5.2.6) was cleaved (Fig. 7).

**Discussion**

It is not unusual that repair enzymes can have a broad substrate range (reviewed in refs. 32, 44, 45). However, it is difficult to understand what the recognition signal(s) is for such enzymes. In this laboratory, it was unexpected that the human APNG could not only cleave εA but with greater efficiency than the original substrate, 3-methyladenine (23). A second recent example is the repair of bulky p-benzoquinone modified dA, dC, and dG by human and Escherichia coli 5'-AP endonucleases (32, 42, 46), inasmuch as the p-benzoquinone moiety adds two additional rings to the bases (47).

In the present work and that of Saparbaev and Laval (35), the two types of mismatches recognized by a single enzyme are εC-G and G/T(U). In the case of G/T(U) mismatches, there is a strong biochemical rationale for their repair because G/T(U) can be produced by deamination of 5-methylcytosine or cytosine, which is estimated to occur with a frequency of ~1.7 × 10^5 events/day/10^10 bases for deamination in rat liver (48). Thus, it would be expected that cells have developed a repair mechanism to ensure survival and genetic integrity.

εC is an environmentally induced modified base but also arises from metabolic processes (3–7). It differs from the effects of deamination of 5-methylcytosine or cytosine in that these latter bases are changed to another normal base, which directly changes the DNA sequence. In contrast, εC is a noninstructive base, which mismatches in replication, both in vitro and in vivo. In vitro, generally adduct-directed mutation frequency is higher than observed in vivo (reviewed in ref. 49). The reason is simple: in vitro experiments measure only base-base interactions in the presence of a purified polymerase, whereas in vivo there are multiple biochemical factors influencing replication. The primary ones are repair and the specificity and multiplicity of DNA polymerases, which are necessary for correction or prevention of miscoding. When qualitatively changed base pairing is the same in vitro and in vivo, it is usually clearer as to which normal bases can be inserted opposite modified bases.

In the case of εC there is a high frequency of pairing with A or T in vitro, in both transcription (8) and replication (10, 17, 18). The work of Singer and Spengler (10), Zhang et al. (17), and Shibutani et al. (18) all agree that there is little or no εC-C or εC-G pairing that occurs. In vivo, by using different systems and assays, the same preferences for εC pairing with A or T were reported by several groups (14–16). Mutation frequency varied but primarily εC → T transitions and εC → A transitions resulted. Lethality was generally high, but of the surviving cells, there were up to 80% mutants (16).

It therefore appears logical that a repair system in the cell is also necessary for the removal of the εC base. Indeed, εC is
efficiently cleaved by the human DNA glycosylase described in this work, as compared with the G/T, even though final proof that multiple activities reside in the same enzyme must await genetic approaches such as a gene knockout. The implication that the eC activity is evolutionarily preserved or results from adaptation also suggests a real in vivo role of the enzyme in counteracting those mutagenic effects resulting from eC.

The finding that one DNA glycosylase acts on two seemingly unrelated substrates, eC-G and G/T, poses an interesting structural problem. Do these two substrates appear the same to the enzyme in terms of adduct structure as well as local perturbation of the nucleic acid? Or can the glycosylate accommodate additional yet unknown substrate structures? If so, what are the limits?

This work was supported by National Institutes of Health Grants CA 47723 and CA 72079 (to B.S.) and was administered by the Lawrence Berkeley National Laboratory under Department of Energy Contract DE-AC03-76SF00098. The authors thank Dr. A. Chenna from this laboratory for synthesizing the oligonucleotides used in this study.