G/T mispairs that arise in the DNA of higher eukaryotes as a result of spontaneous hydrolytic deamination of 5-methylcytosines to thymine must be corrected to G/C pairs. We describe here the purification to apparent homogeneity of the enzyme that initiates this repair process by excising the mispaired thymine from the heteroduplex to generate an apyrimidinic site. The enzymatic activity could be attributed to a 55-kDa polypeptide, which was purified from extracts of HeLa cells by a combination of conventional and DNA-affinity chromatography. The enzyme is a mismatch-specific thymine-DNA N-glycosylase, capable of hydrolyzing the carbon-nitrogen bond between the sugar-phosphate backbone of the DNA and a mispaired thymine. In addition to the G/T, the enzyme can remove thymine also from C/T and T/T mispairs in the order G/T→C/T→T/T. It has no detectable endonucleolytic activity on apyrimidinic sites and does not catalyze the removal of thymine from AT pairs or from single-stranded DNA.

Cytosine residues in the DNA of higher eukaryotes are often methylated. This modification extends to approximately 5% of all cytosines in mammalian DNA (Estor et al., 1984). Although cytosine methylation has been studied for many years, its function in the DNA is not clearly understood. It has been attributed roles in processes ranging from the regulation of gene expression, genomic imprinting, and transcription to the maintenance of genome stability, recombination, and the prevention of excessive homology. X chromosome inactivation (see, for example, Jost and Saluz (1992)), and it would appear that X chromosome inactivation is a necessary function of the X chromosome. In this context, it is important to note that it is involved in most if not all of these. Irrespective of what the actual function(s) of 5-methylcytosine may be at the molecular level, perturbations of the DNA methylation pattern lead to aberrations in cellular differentiation (see, for example, Jones and Taylor (1980)). In a recent series of experiments involving transgenic mice, Estor and colleagues (Li et al., 1992) demonstrated that knocking out the DNA cytosine methylase gene leads to lethal defects during embryonal development, thus providing direct evidence for this hypothesis.

The loss of 5-methylcytosine through deamination results in a change in the DNA methylation pattern, which could be detrimental to the cell. In addition, the C→T mutation resulting from the deamination process in the body of a gene could alter the sequence of the encoded protein, which could have potentially devastating results. Thus, for example, the inactivation of the p53 tumor suppressor protein in a large proportion of bladder carcinomas has been attributed to deamination of 5-methylcytosine (Rideout et al., 1990). In fact, this latter process has long been held responsible for a large proportion of C→T transition mutations, both in prokaryotes (Duncan and Miller 1980) and in higher eukaryotes, where it has been postulated to be the major cause of the conversion of CpG dinucleotides to TpG and CpA (Bird et al., 1979). On the basis of these data, it had been suggested that G/T mismatches associated with the deamination process are not repaired, at least not with high efficiency.

In contrast to these predictions, however, we showed that G/T mismatches are very efficiently corrected to G/C pairs in cultured mammalian cells (Brown and Jiricny, 1987), by a pathway apparently dedicated to this type of lesion and distinct from the general mismatch repair system employed in the correction of non-synonymous errors (Brown and Jiricny, 1988; Modrich, 1991). Later, using nuclear extracts of HeLa cells and synthetic oligonucleotide substrates, we demonstrated that the specific G/T to G/C repair event was initiated by a mismatch-specific thymine-DNA glycosylase (Wibauer and Jiricny, 1990).

We now describe the purification of this activity, a 55-kDa protein, from HeLa cells.

EXPERIMENTAL PROCEDURES

All the reagents and solvents used in this work were of analytical grade purity. HeLa cells were purchased from Computer Cell Culture Center (Mons, Belgium). Phosphocellulose P11 was from Whatman, the other chromatography matrices, DEAE-Sepharose® Fast Flow, Mono-Q®, and HiLoad S-Sepharose® HP fromPharmacia. Streptavidin-de- rivatized Dynabeads were from Dynal AS., Biotin-16-dUTP from Boehringer Mannheim, and [32P]dATP, [Y-32P]dATP, [3H]dTTP, and the Rainbow prestrained protein molecular weight standards were from Amersham Corp. Low molecular weight protein standards were from BioRad. The oligonucleotides were synthesized on an Applied Biosystems model 380B automated synthesizer and purified by polyacrylamide gel electrophoresis. Bovine serum albumin fraction V (BSA) was from Life Technologies, Inc. Standard molecular biological manipulations (labeling of oligonucleotides, ethanol precipitation of DNA, polyacrylamide gel electrophoresis, etc.) were carried out as described by Sambrook et al. (1989).

Synthesis of Affinity Matrix—The G/U affinity matrix was prepared by annealing 14-mer oligonucleotides 5'-GATCCGTGACCCCTG-3' and 5'-GATCCAGTGUACCCG-3' in annealing buffer (10 mM Tris- HCl, pH 8.0, 10 mM MgCl2) as described (Jiricny et al. 1986). 200 μg of the annealed 14-mer duplexes were allowed to ligate end-to-end overnight at 12°C in ligation buffer (25 mM Tris-HCl, pH 7.4, 5 mM MgCl2, 5 mM dithiothreitol, 0.25 mM spermidine, 1.25 mM histamine cobalt chloride, 10 μg/ml BSA), 1 mM ATP, and 10 μl of T4 DNA ligase (400 units/ml, New England Biolabs) in a total volume of 200 μl. After ligation, the DNA was recovered by ethanol precipitation and the dried DNA pellet was dissolved in 48 μl of H2O. The 5'-overhangs were then filled-in with Sequenase version 2.0 (U. S. Biochemical Corp., 2 μl, 13 units/μl stock diluted 1:8 in the presence of 0.3 μM dGTP, 0.3 μM dCTP, 0.15 μM dATP, 0.15 μM dTTP and 0.3 μM dITP, as described by Sambrook et al. (1989).
dATP, 0.15 μM [α-32P]dATP (3000 Ci/mmol), and 0.1 μM Biotin-16-dUTP in a total volume of 20 μl for 10 min at 37 °C. Free nucleotides were removed using spin-column centrifugation with a 1-ml syringe filled with Sephadex G-50 Superfine (Pharmacia). 2.5 ml of Dynabeads M-280 derivatized with streptavidin were preincubated as directed by the manufacturer and preincubated once with HE buffer (25 mM Hepes/NaCl, pH 7.8, 1 mM EDTA, 1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride, 0.5 mM spermidine, and 0.1 mM spermine) at 4 °C for 30 min. They were then homogenized in a glass/glass Dounce homogenizer (Belco) with 20 strokes of a tightly fitting pestle. Glycerol was added to a final concentration of 20% (v/v), followed by a saturated and neutralized (NH₄)₂SO₄ solution (11 ml/100 ml of extract). The mixture was allowed to stand for 30 min, and the extract was cleared by centrifugation in a Beckman ultracentrifuge, using a Ti-70 rotor at 80,000 rpm for 1 h at 4 °C.

The cleared extract was diluted 1:4 with HE buffer and incubated batchwise with DEAE-Sepharose Fast Flow (equilibrated in HE buffer containing 0.1 mM NaCl, 10-15 mg of protein/ml of matrix) for 1 h at 4 °C. The matrix was washed stepwise in a sintered glass funnel with two bed volumes each (collected separately) of HE buffer containing 0.1 and 0.5 M NaCl. The flow-through and the first wash from the DEAE-Sepharose Fast Flow were directly incubated batchwise for 1 h at 4 °C with phosphocellulose P11 (25-30 mg of protein/ml of matrix), which had been preincubated as directed by the manufacturer and equilibrated with HE buffer containing 0.1 mM NaCl. The phosphocellulose was washed stepwise as before with three matrix volumes each (collected in 3 aliquots) of HE buffer containing 0.1, 0.3, and 0.5 mM NaCl.

The flow-through and the first wash from the DEAE-Sepharose Fast Flow were directly incubated batchwise for 1 h at 4 °C with phosphocellulose P11 (25-30 mg of protein/ml of matrix), which had been preincubated as directed by the manufacturer and equilibrated with HE buffer containing 0.1 mM NaCl. The phosphocellulose was washed stepwise as before with three matrix volumes each (collected in 3 aliquots) of HE buffer containing 0.1, 0.3, and 0.5 mM NaCl.

The flow-through and the first wash from the DEAE-Sepharose Fast Flow were directly incubated batchwise for 1 h at 4 °C with phosphocellulose P11 (25-30 mg of protein/ml of matrix), which had been preincubated as directed by the manufacturer and equilibrated with HE buffer containing 0.1 mM NaCl. The phosphocellulose was washed stepwise as before with three matrix volumes each (collected in 3 aliquots) of HE buffer containing 0.1, 0.3, and 0.5 mM NaCl.

Following the phosphocellulose chromatography step, active fractions (0.2 w/v, 0.25, and 0.5 w/v in Fig. 1a) equivalent to 120 g of HELa cells were pooled, diluted 1:2 with HE buffer, and loaded on a HiLoad S-Sepharose High Performance (HP) 26/10 column (HP, Pharmacia) with a peristaltic pump at 4 °C. All following steps were carried out at room temperature. The column was washed with HE buffer containing 0.25 mM NaCl, and the proteins were eluted with a two-step gradient, a 150-ml gradient from 0.25 to 0.5 mM NaCl followed by an 80-ml gradient from 0.5 to 1 mM NaCl.

Active fractions from the HP-S FPLC (28-38, Fig. 1b) were diluted 1:4 with HE buffer and loaded on a Mono Q FPLC column (1 ml), equilibrated with HE buffer containing 0.1 mM NaCl. Proteins were eluted from the column with a 50-ml linear gradient from 0.1 to 0.5 NaCl in HE buffer. Due to the small amount of protein loaded onto the column, the activity eluted as a broad peak. The active fractions were pooled and reloaded onto the same column, eluted with the same buffer gradient.

The active fractions (10-14) from all the Mono-Q columns were pooled, diluted 1:2 with HE buffer in a siliconized 50-ml Falcon tube, and incubated with 1 ml of streptavidin-Sepharose beads (prepared as described above), equilibrated with HE buffer containing 0.1 mM NaCl for 1 h at 4 °C. All magnetic beads were concentrated in one 1.7-ml MultiTak® tube (Multi-Tech Inc.) using the magnetic particle concentrator for microtubes, supplied by the manufacturer. The beads were washed with three 200-μl volumes of HE buffer containing 0.1, 0.2, 0.4, and 0.5 mM NaCl, respectively. Active fractions were pooled, diluted 1:4 with HE buffer in a 4-ml siliconized tube, and incubated with 1.5 ml of streptavidin-Sepharose beads, as described above. All the following steps were performed as described for the first affinity chromatography. Following the washing steps, the active fractions were eluted with a 20-μl volume of the same buffer containing 0.01 mM ZnCl₂, 0.01 mM dithiothreitol, in a total volume of 25 μl at the temperature and for the times indicated in the respective figure legends. Under these conditions, the excision of the mispaired thymine is in most cases accompanied by a cleavage of the labeled strand of the duplex at the 3'-side of the apyrimidinic (AP) site, presumably by a base-catalyzed β-elimination. After the addition of an equal volume of stop buffer (50 mM Tris-HCl, pH 7.5, 25 mM EDTA, 2% glycerol, 5 mM NaCl, 0.1% SDS), the mixture was washed with 1 ml of water, the flow-through and the wash were added, and the proteins were loaded on the gels as described.

**Band Shift Assay**—The band-shift assays were carried out essentially as described previously (Jiricny et al., 1988). 40 fmol of the 5'-AP-labeled 34-mer oligonucleotide were incubated with the relevant protein fraction in 1× binding buffer for 30 min at 37 °C in a final volume of 20 μl, in either the presence (Fig. 2a, b) or absence (Fig. 2c, d) of 400 nM of a competitor (poly(dC)-poly(dG)) (Pharmacia). All of 20 μl of 5% Ficoll were then added, and 5 μl of the reaction mix were loaded on a 7% nondenaturing polyacrylamide gel made in TAE buffer (4 mM Tris acetate, pH 7.5, 1 mM EDTA). Electrophoresis was carried out at 10 V/cm for 70 min.

**G/T 90-mer Duplex**—This substrate consisted of a 90-mer G ongucleotide annealed with a 90-mer T oligonucleotide (see Scheme 1), where the mispaired thymine was labeled to a high specific activity with 3H. It was constructed as described earlier (Wiebauer and Jiricny, 1990). The specific activity of the duplex was 48.7 Ci/mmol.

**Specific Activity Measurements**—In order to measure the specific activity of the enzyme preparations, the active fractions listed in Table 1 were incubated in a mix containing 800 fmol of the 96-mer G[3H]T in binding buffer and 20 μl of whole cell extract, DEAE-Sepharose, HP-S, and Mono-Q FPLC fractions or 4 μl of the 0.4/1 affinity column fraction (see Fig. 4a) in a total volume of 100 μl. 25-ml aliquots were removed after 10 min, 30 min, and 17 h and mixed with an equal volume of stop buffer, and the reaction was terminated by incubation at 37 °C for 20 min. The mixtures were diluted to 300 μl with water and applied directly on a 0.3-ml DEAE-Sepharose Fast Flow column. The column was washed with 1 ml of water, the flow-through and the wash were combined, 5 μl of ReadySafe™ liquid scintillation mixture (Beckman) were added, and the mixture was counted for 5 min. As the oligonucleotides and free nucleotides remain bound to the ion-exchange column (data not shown), the amount of radioactivity contained in the flow-through represents the amount of thymine liberated by the glycosylase (see Section 3).

**Renaturation of the Glycosylase Activity**—The proteins were electrophoresed on discontinuous SDS-polyacrylamide gels (5% stacking gel, 10% separating gel) as described by Sambrook et al. (1989). The gel was loaded as follows: lanes 1 and 3, 10 μl of prestained Rainbow protein molecular weight marker (Amersham Corp.); lanes 2, 10 μl of fraction 2/4 (Fig. 4a); lane 4, 8 μl of Rainbow marker (diluted 1:100); lane 5, 2 μl of fraction 0/4. After electrophoresis, lane 2 was cut into twelve 0.4-mm slices, and lanes 4 and 5 were silver-stained (Fig. 5a). The proteins were eluted from the gel slices overnight at 37 °C with vigorous shaking in 400 μl of elution buffer (1× HE buffer, 10 mM NaN, 2 mM dithiothreitol, 0.1% SDS, 0.1 mg/ml BSA) and then precipitated with 4 volumes of ice-cold acetone. The dried protein pellet was dissolved in 25 μl of guanidinium buffer (1× HE buffer, 50 mM NaCl, 2 mM dithiothreitol, 0.1 mg/ml BSA, 6 M guanidinium HCl) and denatured for 30 min at room temperature. Renaturation was performed at 4 °C essentially as described by Hao and Georgopoulos (1980); the protein solution was diluted 50-fold (1.25 ml) with the dilution buffer (as above but without guanidinium HCl). The fractions were concentrated to 40 μl in a Centricon 30 microconcentrator (Amicon Corp., prewashed with 0.5 ml of dilution buffer containing 0.5 mg/ml BSA). 10 μl of the concentrated fractions were used in the enzyme activity assay as described above.

**AP Endonuclease Activity Assays**—To test the enzyme preparations for the presence of AP endonuclease activity, the nicking assays were carried out at pH 6.8. Following digestion with proteinase K and ethanol precipitation, the samples were either resuspended in 4 μl of binding buffer (pH 8.6) or in 1× binding buffer (pH 7.5), boiled at 90 °C for 30 min. 4 μl of urea dye were then added to both assays, and...
the samples were loaded on a denaturing polyacrylamide gel made and run in 1× BBE buffer (90 mM Bis-Tris borate, pH 6.8, 2 mM EDTA).

**RESULTS**

**Purification of the Mismatch-specific Thymine-DNA Glycosylase**—During the purification, the enzymatic activity was monitored by the nicking assay. The specific activities of the respective fractions were estimated from the amounts of thymine liberated from the 90-mer oligonucleotide G/[3H]T (see “Experimental Procedures”).

The starting material for the purification were whole cell extracts from 900 g of HeLa cells. These were used because, although the protein was found predominantly in the nuclear fraction in fresh cell extracts, in extracts from commercially-available frozen cells, up to 50% of our activity appeared in the cytoplasmic fraction (data not shown).

The initial step in the purification scheme was a batchwise “filtration” chromatography of the extract on an anion-exchanger (DEAE-Sepharose Fast Flow), which removed approximately one half of the total proteins and most of the nucleic acids (Table I). As can be seen in Fig. 1a, the G/T-processing enzyme was found in the flow-through and in the 0.1 M NaCl washes. Cation-exchange chromatography of the pooled active fractions on Phosphocellulose P11 afforded a 4.7-fold enrichment with 0.3-0.5 M salt (Fig. 1a). The following step, FPLC on an anion-exchanger (DEAE-Sepharose Fast Flow), which removed approximately 79,800 units/mg, 3.0% recovery, due to the presence of nonspecific competitors and/or inhibitors in this fraction.

### Table I

**Purification of thymine-DNA glycosylase from 900 g of HeLa cells**

| Fraction       | Protein (mg) | Specific activitya | Recovery (%) |
|----------------|--------------|--------------------|--------------|
| WCE            | 79,800       | ND                 | ND           |
| DEAE-Sepharose | 42,622       | 85                 | 100          |
| Phosphocellulose| 2,610        | 400                | 29           |
| Hp-S           | 218          | 490                | 3.0          |
| Mono-Q         | 7.2          | 4,100              | 0.8          |
| DNA-Affinity 1 | 0.334        | 102,940            | 0.1          |
| DNA-Affinity 2 | 0.011        | ND                 | NDb          |

a One unit is defined as 1 fmol of liberated [3H]thymine after 10-min incubation at 37 °C.

b ND, not determined.

The specific activity of the whole cell extract was not used as the 100% reference point in the calculation of the recovery yield, due to the presence of nonspecific competitors and/or inhibitors in this fraction.

With our activity. In particular, we tried to ensure that we could separate our activity from uracil-DNA glycosylase, which is present in the HeLa extracts in considerable amounts. For this reason, we routinely monitored both the thymine- and the uracil-DNA glycosylases by a nicking assay, using G/T and G/U duplexes, as well as a single-stranded uracil-containing oligonucleotide (data not shown). The results of these assays indicated that the two activities co-purified throughout the various purification procedures. However, we noted that the thymine-DNA glycosylase appeared to be a significantly slower acting enzyme than the uracil glycosylase, as witnessed by the long incubation times needed to visualize enzymatic activity in the nicking assays. We postulated that our enzyme may thus form protein-DNA complexes that are sufficiently long-lived to enable us to use a small (1-ml) Mono-Q FPLC column. This step yielded the G/T-processing activity in a very concentrated form in a total volume of 15 ml. The elution profiles of the HP-S and Mono-Q FPLC columns are shown in Fig. 1 (b and c, respectively). The protein profiles of the activity-containing fractions are shown in Fig. 3.

### Diagram showing the synthetic oligonucleotide substrates used in this study. The 34-mer corresponds to the fragment of the 90-mer designated by dashed lines. N, G, A, T, or C. N', G', C', or T. The solid arrows designate the sites of cleavage of a G/C duplex by the three restriction endonucleases HindII, AccI, and SalI. 5'-End labeling of the C strand of the 34-mer G/C duplex, followed by the restriction digest using these three enzymes, generated the size marker shown in Fig. 4c, 5b, and 6 (center lane). 5'-End labeling of the G strand of the 34-mer G/C duplex, followed by similar restriction digests, generated the size marker shown in Fig. 6 (right lane).
The protein that could form relatively stable complexes with G/T but not A/T or G/C heteroduplexes. The same fractions also gave a band-shift with a G/U duplex (Fig. 2). We discounted the possibility that this latter complex was due to uracil-DNA glycosylase, as no band-shift could be seen with the same GN oligonucleotide incubated with a bacterial uracil-DNA glycosylase, as no band-shift could be seen with the same GN oligonucleotide incubated with a bacterial uracil-DNA glycosylase. 2

As the G/U-containing complex had similar electrophoretic mobility as that containing the G/T oligonucleotide (Fig. 2), we concluded that it was formed between the thymine-DNA glycosylase and the G/T or the G/U duplex, respectively. Due to the fact that the binding to the G/U substrate appeared noticeably stronger, it was this latter substrate that we decided to employ as our affinity matrix. Following two rounds of DNA affinity purification, only a single protein band was visible by silver staining on a SDS-polyacrylamide gel. It migrated with an apparent molecular mass of 55 kDa (Fig. 3, lanes A1.1 and A1.2).

The 55-kDa Protein Is a Thymine-DNA Glycosylase—In order to ensure that the protein band eluting from the affinity matrix with 0.4–0.5 M NaCl represented the thymine-DNA glycosylase, we carried out a series of experiments shown in Fig. 4. In panel a we show the protein band pattern eluting from the affinity matrix. It is evident that the 55-kDa protein appears in the 0.4 and 0.5 M NaCl fractions. Panel b shows a band-shift experiment, where the latter two fractions can be seen to form protein/DNA complexes with a G/T mismatch-containing oligonucleotide duplex. This same substrate is shown to have been nicked at the site of the mispair upon incubation with the 0.4 and 0.5 M NaCl fractions (panel c).
Additional evidence as to the identity of the G/T mismatch processing activity and the 55-kDa protein band comes from an elution/renaturation experiment shown in Fig. 5, which shows that the nicking activity and the 55-kDa protein band co-migrate in a denaturing SDS-polyacrylamide gel. These experiments thus provide convincing evidence that the 55-kDa protein represents the thymine-DNA glycosylase.

**Substrate Specificity of the Thymine-DNA Glycosylase**—As shown above (see also Wiebauer and Jiricny (1990)), the enzyme catalyzes the removal of a mispaired thymine from G/T mismatches. We wanted to test whether its activity was restricted solely to this mispair or whether the enzyme was able to act also on other thymine-containing mismatches. To this end we constructed 34-mer heteroduplexes containing G/C, G/T, A/T, T/T, and C/T base pairs and incubated them with the purified glycosylase preparation (fraction 0.4/1, Fig. 4a). As can be seen from Fig. 6, no processing of the Watson-Crick (G/C and A/T) duplexes was observed, but the thymine-containing heteroduplexes were seen to be nicked at the site of the mispair, with the nicking efficiency decreasing in the order G/T >> C/T > T/T. It is unlikely that the latter two mismatches represent true substrates in vivo. Rather, it would seem more probable that under the conditions of the assay, i.e., where only the purified protein and the mispaired oligonucleotide are present, even a small structural deviation from a Watson-Crick base pair is sufficient for recognition by the enzyme.

The Thymine-DNA Glycosylase Lacks an Associated AP Endonuclease Activity—Our initial studies with HeLa nuclear extracts (Wiebauer and Jiricny, 1989) suggested that following the action of the thymine-DNA glycosylase, the baseless sugar-phosphate residue was removed by a two-step excision process, which first “nicked” the DNA 3'- from the AP site by β-elimination and then removed the baseless sugar-phosphate by 3'→5' exonuclease. As this mechanism of AP site processing is normally associated solely with AP lyases, i.e., enzymes possessing both a glycosylase and an AP endonuclease activity (see Weiss and Grossman (1987) for review), we decided to test whether this latter function was also associated with the purified thymine-DNA glycosylase preparations. Incubation of the G/T oligonucleotide with an active fraction, followed by separation of the fragments by a conventional denaturing polyacrylamide gel electrophoresis in 1 X TBE buffer, always resulted in the “nicking” of the substrate. This would suggest that protein fractions contained, in addition to the glycosylase, an AP endonuclease. However, as bacterial uracil-DNA glycosylase, which is known to possess no AP endonuclease activity, could also be seen to produce a nick in a similar, G/U-containing substrate (data not shown), we suspected that the cleavage of the sugar-phosphate backbone at the apyrimidinic site was an artifact of our analytical system. We thus tested the possibility that the observed β-elimination reaction was catalyzed by the conditions of the assay: high pH (8.3) and high temperatures (heating of sample in loading dye for 5 min at 95 °C). We therefore repeated the nicking assay experiments at a pH below 7, i.e., under conditions that do not favor β-elimination. Indeed, as shown in Fig. 7 (lanes -NaOH), the labeled T strand of the G/T duplex remained mostly intact following incubation with the respective protein fractions and electrophoresis. The same substrates could be shown to contain AP sites by treatment with 0.1 N NaOH (Fig. 7, lanes +NaOH), which leads to a quantitative double β-elimination (2′-3′ and 4′-5′) reaction that first cleaves the DNA 3′ from the dephosphorylated nucleotide and then removes the baseless sugar to leave a phosphate at the 3′-end of the labeled fragment (Maxam and Gilbert, 1977). This experiment therefore demonstrates that the purified mismatch-specific thymine-DNA glycosylase contains no intrinsic AP endonuclease activity.

**DISCUSSION**

In nuclear extracts from HeLa cells, a G/T mispair incorporated in a synthetic oligonucleotide duplex is addressed initially by a mismatch-specific thymine-DNA glycosylase, which excises the mispaired thymine to generate an apyrimidinic site opposite the guanine (Wiebauer and Jiricny, 1990). The exact mechanism of processing of this AP site was not clear. Our initial in vitro studies (Wiebauer and Jiricny, 1989) suggested that the DNA backbone was first cleaved at its 3′-side by a process of β-elimination, following which a 3′→5′ exonucleolytic step generated a single nucleotide gap, which was, in turn, filled in with a dCMP residue by polymerase-β to yield a G/C base pair. The remaining nick was then sealed by a DNA ligase (Wiebauer and Jiricny, 1990). Our present findings suggest that the β-elimination at the AP site was an artifact of our assay system or that it may have been catalyzed by an unknown basic factor(s) present in the extracts, as the purified
The 55-kDa protein is the thymine-DNA glycosylase. Renaturation of the protein eluted from slice 7 of the SDS-polyacrylamide minigel shown in panel a led to the recovery of the G/T nicking activity (panel b, lane 7). a, silver-stained half of a denaturing 10% SDS-polyacrylamide Minigel; left lane, 8 μ1 of prestained Rainbow protein molecular weight marker (Amersham) (myosin, 200 kDa; phosphorylase b, 92.5 kDa; BSA, 69.0 kDa; ovalbumin, 46 kDa; carbonic anhydrase, 30.0 kDa; trypsin inhibitor, 21.5 kDa; lysozyme, 14.3 kDa) diluted 1:100; right lane, 2 μ1 of fraction OM1 from the first affinity column (see Fig. 4a); the positions and numbering of the 4-mm slices obtained from the other, symmetrically loaded half of the gel (see "Experimental Procedures") are shown on the right. b, nicking activity of the proteins eluted from the gel slices and renatured (see "Experimental Procedures") were assayed by incubation with the 34-mer oligonucleotide (labeled at the 5'-end of the T strand with 32P) and digested with HincII, AccI, and SalI.

Thymine-DNA glycosylase does not possess any detectable endonucleolytic activity (Fig. 7).

The G/T repair process thus closely resembles the classical base-excision repair pathway (Dianov et al., 1992) (see Weiss and Grossman (1987) and Lindahl (1993) for reviews). The concept of a glycosylase excising an unmodified DNA base was initially rather surprising, given that all DNA glycosylases characterized until recently were restricted in their substrate specificity to modified or damaged DNA bases (Lindahl, 1982; Sancar and Sancar, 1988). It now appears, however, that mismatch-specific DNA glycosylases may be limited neither to the G/T mispair nor to mammals. Other organisms have also evolved enzymes capable of acting independently of the replication-associated mismatch repair machinery. Thus the MutY protein of Escherichia coli was shown to catalyze the excision of adenine from G/A and, to a lesser extent, A/C mispairs (Au et al., 1989; Tsai-Wu et al., 1992), and a similar activity was also described in mammalian cells (Yeh et al., 1991). Clearly, such glycosylases must have different requirements for substrate recognition, in that only bases in a mispair may be removed. This is a necessary constraint on the enzyme, which prevents the loss of natural bases from Watson-Crick base pairs or

---

**Fig. 6.** The glycosylase catalyzes the removal of thymine from all thymine-containing mispairs. 40 fmol of each 34-mer oligonucleotide duplex (G/C*, G/T*, A/T*, T/T*, C/T*, T*/G, T*/T, and T*/C) were incubated for 30 min at 37 °C with 1 μl of the active affinity fraction 0.4/1 (Fig. 4a) as described under "Experimental Procedures." The asterisk indicates the 32P-labeled strand. M, marker oligonucleotide G/C, labeled as denoted by asterisk and digested with HincII, AccI, and SalI. The figure shows an autoradiogram of a denaturing 20% polyacrylamide gel. The band migrating at the height of the SalI or AccI marker in the G/T and T*/G lanes, respectively, represents a contaminating 33-mer in the oligonucleotide preparation, visible due to the gross overexposure of the autoradiogram.

**Fig. 7.** The thymine-DNA glycosylase possesses no intrinsic AP endonuclease activity. 40 fmol of each 90-mer oligonucleotide G/T (labeled at the 5'-end of the T strand with 32P) were incubated for 1 h at 37 °C at pH 6.8 with the active fractions from the Mono-Q (1.9 μg of protein) and DNA affinity (0.5 μg of protein) stages of purification. With no NaOH treatment, only a small amount of cleavage of the labeled strand was observed (lanes +NaOH), which is due to the spontaneous β-elimination at AP sites. Treatment with 0.1 x NaOH resulted in the quantitative cleavage at these sites (lanes +NaOH). Q and Af, active fractions from the Mono-Q FPLC and the first DNA-affinity chromatography purification stage. (The latter (+NaOH) reaction product is shorter than the former (-NaOH) oligonucleotide by 1 baseless sugar residue and has an additional charge on the 3'-terminal phosphate. It therefore migrates faster in polyacrylamide gels.)
single-stranded DNA.

We proposed (Wiebauer and Jiricny, 1989) that the biological role of the thymine glycosylase is the correction of G/T mispairs arising from hydrolytic deamination of 5-methylcytosine. In the light of the above data, the enzyme apparently satisfies all the criteria required for this function, with the possible exception of one; SV40 transfection experiments (Brown and Jiricny, 1987) showed that it lacked specificity for G/T mismatches in the context of CpG dinucleotides, the sites of mammalian cytosine methylation. The thymine-DNA glycosylase could potentially address also G/T mispairs arising as biosynthetic errors and might thus interfere with the proper functioning of the replication-associated mismatch correction process by ignoring the strand bias required in the repair of biosynthetic errors (Modrich, 1991). However, a recent report by Ullah and Day (1993) suggests that the enzyme may have a preference for CpG-associated mispairs in vitro, which would further substantiate our hypothesis that its biological role is the correction of deamination-associated G/T mispairs.

One should not, however, discount the possibility of a competition between the replication-associated and the thymine glycosylase-mediated mismatch repair processes. Transfection experiments with mismatch-carrying SV40 heteroduplexes clearly showed that, unlike all other mismatches, the G/T mispairs were predominantly repaired by the latter pathway, but that a small fraction of the transfected G/T heteroduplexes (approximately 8%) appeared to have been addressed by a different mismatch correction system, which lacked the G/T → G/C directionality (Brown and Jiricny, 1987). We postulated (Jiricny, 1991) that these molecules may have contained random nicks in the circular DNA and were therefore addressed by the nick-directed long-patch mismatch repair pathway (Holmes et al., 1990; Thomas et al., 1991). These data would imply that these two mismatch correction systems can indeed compete for the same substrate. It could be argued, however, that both the SV40 transfection experiments and the in vitro mismatch correction assays represent artificial systems that may or may not mirror the situation in vivo. We must not discount the possibility, for example, that in a cell the two pathways are temporally compartmentalized, in that they may not act in the same stage of the cell cycle. The recently reported isolation of a mutator cell line lacking a G/T-binding protein (Branch et al., 1993), which is presumably a component of the replication-associated mismatch repair process (Hughes and Jiricny, 1992) ought to help us provide the answer to these questions.

Acknowledgments—We express our gratitude to Tomas Lindahl for the gift of the bacterial uracil-DNA glycosylase, to Karin Wiebauer, Melyn Hughes, and Raffaele De Francesco for numerous helpful discussions, to Ricardo Cortese, Paolo Gallinari, Paolo Monaci, and Andrew Wallace for critical reading of the manuscript, to Philippe Neuner for the oligonucleotides, and to Yves Cully for photographic assistance.

REFERENCES

Au, K. G., Clark, S., Miller, J. H., and Modrich, P. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 8877-8881
Baxter, T., Hellewell, S. B., and Ingram, V. M. (1994) Mol. Cell. Biol. 4, 1800-1806
Bird, A., Taggart, M., and Smith, B. (1979) Cell 17, 889-901
Branch, P., Aquilina, G., Bignami, M., and Karran, P. (1993) Nature 362, 652-654
Brown, T. C., and Jiricny, J. (1987) Cell 50, 945-950
Brown, T. C., and Jiricny, J. (1988) Cell 54, 765-771
Dianov, G., Price, A., and Lindahl, T. (1992) Mol. Cell. Biol. 12, 1605-1612
Duncan, B. K., and Miller, J. H. (1980) Nature 287, 560-563
Hager, D. A., and Burgess, R. R. (1980) Anal. Biochem. 106, 76-86
Holmes, J., Clerk, S., and Modrich, P. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 5837-5841
Hughes, M. J., and Jiricny, J. (1992) J. Biol. Chem. 267, 23876-23882
Jiricny, J. (1991) in Nucleic Acids and Molecular Biology (Eckstein, F., and Lilly, D. M., eds) pp. 72-83, Springer-Verlag, Berlin
Jiricny, J., Wood, S. G., Martin, D., and Ubasawa, A. (1986) Nucleic Acids Res. 14, 6579-6590
Jiricny, J., Hughes, M. J., Cormann, N., and Rudkin, B. B. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 8860-8864
Jones, P. A., and Taylor, S. M. (1989) Cell 20, 85-93
Jost, J.-P., and Saluz, H. P. (1993) DNA Methylation: Molecular Biology and Biological Significance, Birkhauser-Verlag, Basel
Li, E., Bestor, T. H., and Jaenisch, R. (1993) Cell 76, 915-926
Lindahl, T. (1982) Annu. Rev. Biochem. 51, 61-87
Lindahl, T. (1993) Nature 362, 709-715
Maxam, A. M., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 560-564
Modrich, P. (1991) Annu. Rev. Genet. 25, 229-252
Rideout, W. M., Coetsee, G. A., and Olumi, A. F. (1990) Science 249, 1288-1290
Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Sancar, A., and Sancar, G. B. (1988) Annu. Rev. Biochem. 57, 29-57
Thomas, D. C., Roberts, J. D., and Kunkel, T. A. (1991) J. Biol. Chem. 266, 3744-3751
Tsui-Wu, J.-J., Liu, H.-F., and Lu, A.-L. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 8779-8783
Ullah, S., and Day, R. S. (1993) Nucleic Acids Res. 21, 1281-1287
Weiss, B., and Grossman, L. (1987) Adv. Enzymol. 59, 3-43
Wiebauer, K., and Jiricny, J. (1988) Nature 335, 234-236
Wiebauer, K., and Jiricny, J. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 5642-5645
Yeh, Y.-C., Chang, D.-Y., Masin, J., Lu, A.-L. (1991) J. Biol. Chem. 266, 6480-6484