Measurement of deaminated cytosine adducts in DNA using a novel hybrid thymine DNA glycosylase

Chia Wei Hsu[1,2] , Mark L. Sowers[1,2] , Tuvshintugs Baljinnyam[1], Jason L. Herring[1], Linda C. Hackfeld[1], Hui Tang[1], Kangling Zhang[1], and Lawrence C. Sowers[1,3,*]

From the [1]Department of Pharmacology and Toxicology, [2]MD-PhD Combined Degree Program, and [3]Department of Internal Medicine, University of Texas Medical Branch, Galveston, Texas, USA

Edited by F. Peter Guengerich

The hydrolytic deamination of cytosine and 5-methylcytosine drives many of the transition mutations observed in human cancer. The deamination-induced mutagenic intermediates include either uracil or thymine adducts mispaired with guanine. While a substantial array of methods exist to measure other types of DNA adducts, the cytosine deamination adducts pose unusual analytical problems, and adequate methods to measure them have not yet been developed. We describe here a novel hybrid thymine DNA glycosylase (TDG) that is comprised of a 29-amino acid sequence from human TDG linked to the catalytic domain of a thymine glycosylase found in an archaeal thermophilic bacterium. Using defined-sequence oligonucleotides, we show that hybrid TDG has robust mispair-selective activity against deaminated U:G and T:G mispairs. We have further developed a method for separating glycosylase-released free bases from oligonucleotides and DNA followed by GC–MS/MS quantification. Using this approach, we have measured for the first time the levels of total uracil, U:G, and T:G pairs in calf thymus DNA. The method presented here will allow the measurement of the formation, persistence, and repair of a biologically important class of deaminated cytosine adducts.

Cytosine to thymine transition mutations are the most abundant single-base changes observed in human cancer cells (1–5). These mutations are believed to arise from the hydrolytic deamination of cytosine and cytosine analogs, which involves a water molecule adding to the C4 carbon, displacing the amino group (6–10). This reaction results in a mispaired intermediate with guanine (Fig. 1). The deaminated bases comprise an important class of DNA adducts; however, they cannot be measured by current approaches. Methods are therefore needed to measure the formation and persistence of the deaminated mispairs.

Several laboratories have developed sensitive and specific methods for measuring a wide array of DNA base adducts; however, such methods would require either enzymatic or acid hydrolysis prior to analysis (11–16). The mutagenic significance of the deaminated cytosine adducts is a consequence of residing in a base mispair with guanine. Methods that use

DNA hydrolysis eliminate this base-pairing context. Furthermore, PCR-based analytical methods would convert the mispaired intermediate to a G:C base pair and an A:T mutation, erasing the initial mispair context as well.

Other laboratories have used DNA repair glycosylases to selectively remove damaged bases from DNA for analysis by mass spectrometry (MS)-based methods (17–21). Uracil-DNA glycosylase (UDG or UNG) has been used to measure total uracil in DNA; however, UDG removes uracil from single-stranded DNA as well as U:A and U:G base pairs and therefore cannot distinguish a deaminated base pair (U:G) from a dUTP misincorporation event (U:A). On the other hand, thymine DNA glycosylases (TDGs) can remove uracil and thymine selectively from mispairs (U:G and T:G). However, the activity of the human TDG (hTDG) is very weak against T:G (22, 23).

Our previous studies suggested that a thermophile TDG glycosylase from *Methanobacterium thermoautotrophicum* (MIG) might have the requisite selectivity for U and T mispaired with guanine (24). We therefore generated a hybrid TDG (hyTDG) by joining a 29 amino acid sequence shown to substantially increase the activity of hTDG (25) to the catalytic core of MIG. In this article, we describe the cloning and purification of hyTDG. We demonstrate that hyTDG has the requisite selectivity for mispaired intermediates and sufficient activity to release deaminated cytosine analogs from DNA for subsequent analysis using MS-based methods.

Results

Construction and characterization of a hybrid human-thermophile mispaired TDG (hyTDG)

A DNA sequence was constructed containing a His-tag, a sequence encoding a 29 amino acid sequence derived from the amino terminus of the hTDG (25), and the catalytic core of a thermophile TDG, MIG (26–28). The amino acid sequence is shown in Figure 2, and the DNA sequence is shown in Figure S1.

The plasmid encoding this sequence was cloned into BL3 competent cells and induced. Proteins isolated from the cell extract were fractionated, and the His-tagged protein was isolated using a Ni^{2+} column. Isolated protein was analyzed by

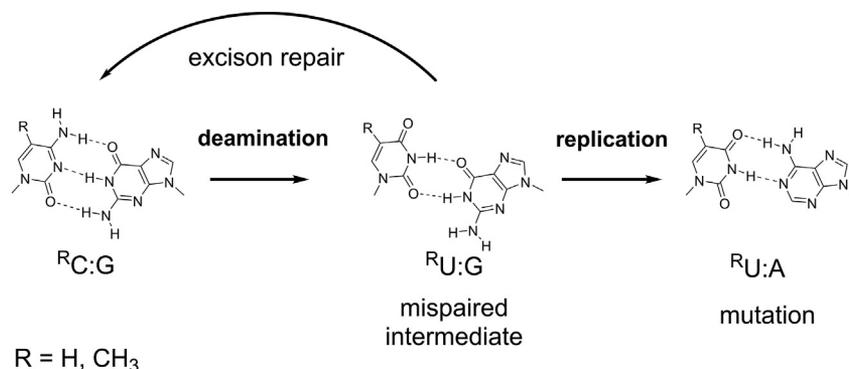* For correspondence: Lawrence C. Sowers, lasowers@UTMB.edu.

Figure 1. Pathway for mutations induced by deamination of cytosine and analogs. Deaminated intermediates can be converted to mutation by DNA replication or repair synthesis. Mispaired intermediates can also be repaired by excision repair pathways.

gel electrophoresis. The predominant band had an apparent molecular weight of 26.5 kDa (Fig. S2).

The purified protein was characterized by LC–MS/MS proteomic methods. A list of observed peptide fragments is provided in Table S1. Protein characterization by MS analysis of tryptic fragments does not always provide 100% coverage. In the case of hyTDG, 98% coverage was obtained. It is valuable to determine which peptides are observed and to map these peptides to regions of the protein critical for enzymatic activity. When hybrid proteins are constructed, analysis of tryptic fragments can be used to confirm the amino acid sequence of the construct.

In the case of hyTDG prepared here, the peptide KVDR/ LDDATNK (amino acids 34–44) was observed (Fig. 3). This peptide is the junction between the human sequence and the MIG enzyme. The peptide SKEKQEKITDTFK (amino acids 18–30) is derived from the human 29 amino acid sequence (Fig. S3). The observation of these peptides confirms correct construction of the hybrid protein.

Previous studies have established functionally important motifs within MIG (26–28), and several peptides mapping to these regions were observed. Single amino acid substitutions within key protein regions can place a glycosylase into a different family, and other changes can dramatically alter substrate selection and enzymatic activity.

The glycosylases that function in base excision repair generally flip a target base from the duplex into an active site pocket. We observed the peptide DPYVILITEILLR**R** (amino acids 71–84), which contains the "**R**" base flipper for this class of glycosylase (Fig. S4).

Among the DNA repair enzymes, the catalytic core of MIG is within a group defined by an integral helix–hairpin–

helix (HhH) motif. We observed the peptide KAILDLPGVGK (amino acids 152–162; Fig. S5), which contains the LPGVGKY HhH motif. The HhH motif consists of two α-helices flanking a β-hairpin with the conserved sequence LPGVGX(K/S), which binds the DNA backbone nonspecifically (28). The HhH motif places the thermophile TDG (MIG) in the same class with the Mut Y and Endo III glycosylases.

A catalytically important conserved aspartic acid residue is common to MIG, Mut Y, Endo III, and glycosylases, which interacts with the C1′ position of the target 2′-deoxyribose. Peptide KAAMV**D**ANFVR (amino acids 176–186; Fig. S6) was observed, and this peptide contains the conserved aspartic acid residue (bold). Also common to MIG, Mut Y and Endo III are an iron–sulfur cluster. We observed the peptide DFNLGLMDFSAII**C**APR (amino acids 221–237; Fig. S7), which contains the first cysteine residue of the iron–sulfur cluster.

Examination of the activity of hyTDG using a gel cleavage assay

We first examined the activity of hyTDG on a series of 5′-6-carboxyfluorescein (6FAM)-labeled oligonucleotides (2.5 pmol) containing single-stranded U or U paired with A or G (Fig. S8). As shown in Figure 4, UDG (2.5 units; 1.6 pmol) cleaves completely single-stranded oligonucleotides containing U as well as duplexes containing U:A base pairs or U:G mispairs in 1 h at 37 °C. In contrast, UDG does not cleave T when paired with A or mispaired with G. Our hyTDG (16.8 pmol) cleaves oligonucleotides containing U:G (95%) or T:G (86%) mispairs but not those containing U:A or T:A base pairs or single-stranded oligonucleotides containing U, in 1 h at 65 °C. The hTDG (29–31) (31.0 pmol) was mispair specific as was hyTDG, cutting U:G (73%) but much less T:G (6%). These data confirm that hyTDG is specific for mispaired pyrimidines and in contrast to hTDG, has substantial activity against T:G mispairs.

The results shown in Figure 4 were obtained with a molar excess of hyTDG. We also conducted a gel assay where 0.5 equivalents of hyTDG were incubated with 25 pmol of the U:G duplex (Fig. S9). Approximately 50% of the duplex was

```
MGHHHHHHSK KSGKSAKSKE KQEKITDTFK VKRKVDRLDD ATNKKRKVFV
STILTFWNTD RRDFPWRHTR DPYVILITEI LLRRTTAGHV KKIYDKFFVK
YKCFEDILKT PKSEIAKDIK EIGLSNQRAE QLKELARVVI NDYGGRVPRN
RKAILDLPGV GKYTCAAVMC LAFGKKAAMV DANFVRVINR YFGGSYENLN
YNHKALWELA ETLVPGGKCR DFNLGLMDFS AIICAPRKPK CEKCGMSKLC
SYYEKCST
```

Figure 2. Amino acid sequence of hyTDG. The 29 amino acid peptide derived from hTDG is shown in *red* and functionally significant peptides observed by mass spectrometry are shown in *bold*. hTDG, human thymine DNA glycosylase; hyTDG, hybrid thymine DNA glycosylase.
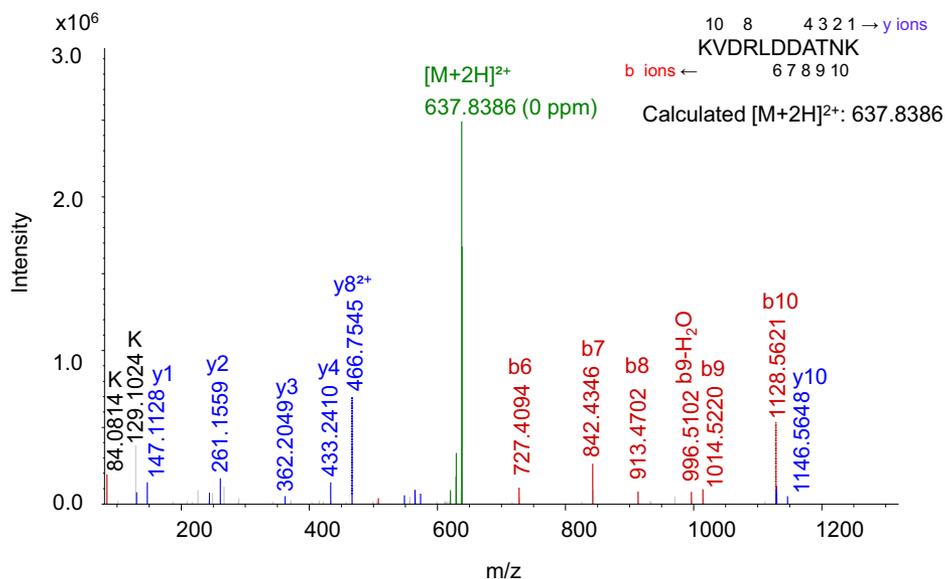
Figure 3. Mass spectrum of peptide KVDR/LDDATNK, which forms the junction between hTDG and MIG sequences. The sequence of the peptide is determined from examination of the b or y ion fragments as indicated in the figure. hTDG, human thymine DNA glycosylase; MIG, thymine DNA glycosylase from *Methanobacterium thermoautotrophicum*.

cleaved at 1 h, and further cleavage was not observed at 4 h. These data suggest that hyTDG excises a target base, remains bound to the abasic site, and does not turn over under the conditions described here.

Examination of the activity of hyTDG using a real-time fluorescence assay

We next analyzed hyTDG cleavage using a real-time fluorescence assay (32, 33) with 5′-6FAM oligos duplexed with a complementary strand containing a 3′-black hole fluorescence quencher 1 (BHQ1) quencher (Fig. S8). In this assay, glycosylase cleavage generates an abasic site, which is then cleaved chemically using *N,N*-dimethylethylenediamine (DMDA) (34), separating the 5′-6FAM from the quencher and allowing continuous monitoring of the fluorescence intensity (Fig. 5).

The real-time assay contained 25 pmol of labeled oligonucleotide duplexes containing either a U:G or a T:G mispair and 25 pmol hyTDG. Reactions were conducted at 65 °C in a 96-well plate in a Roche 480 quantitative PCR instrument. Fluorescence was measured every 20 s. Three independent assays were conducted simultaneously. Fluorescence data were normalized to the total percentage of cleaved oligonucleotide. This was measured by gel electrophoresis after the last fluorescence time point was acquired. Example data are shown in Figure S10. In panel *A*, the raw cleavage data from three independent experiments are shown. In panel *B*, the average of the three runs is shown as a *solid line*, and the standard deviation is represented by *vertical lines*. Finally, a single exponential was fit to the experimental data as shown in Figure S10*C*. The equation describing the theoretical line is

$$Y_t = A(1 - e^{-kt}) \qquad (1)$$

where Y_t is percent cleavage at time t (min), A is the percent cleavage at 60 min, and k is a rate constant (min^{-1}).

Data for the cleavage of U:G-containing and T:G-containing duplexes are shown in Figure 5*A*, and values for A and k (Equation 1) are presented in the figure legend. The initial cleavage velocity, v_o, was calculated from the fitted rate constant and Equation 1 at 5% cleavage. The initial velocity for hyTDG cleavage of the U:G oligonucleotide was measured to be 2.62 ± 0.3 pmol/min and for the T:G oligonucleotide, 1.69 ± 0.20 pmol/min. No cleavage was observed for U:A-containing or T:A-containing oligonucleotides.

In a second experiment, hyTDG cleavage was measured in the presence of 20 μg calf thymus DNA (Fig. 5*B*). The initial cleavage velocities under these conditions were measured to be 2.77 ± 0.40 pmol/min for U:G and 1.49 ± 0.30 pmol/min for T:G. hyTDG cleaves U:G faster than T:G under both conditions. Surprisingly, v_o did not decrease significantly in the presence of excess calf thymus DNA for either U:G or T:G mispairs.

We also examined the activity of MIG on the U:G substrate in this assay (Fig. S11). The v_o for MIG in the absence of calf thymus DNA was measured to be 2.23 ± 0.3 pmol/min. In the presence of calf thymus DNA, it was 1.15 ± 0.2 pmol/min. Our hyTDG is slightly faster than MIG in the absence of excess calf thymus DNA. Furthermore, the rate of hyTDG does not diminish in the presence of excess calf thymus DNA, unlike MIG. These data suggest that the major advantage of the appended 29 amino acid sequence from hTDG is to facilitate scanning of the DNA for a target base.

Examination of pyrimidines released from oligonucleotides and DNA by hyTDG

The aforementioned assays allow the examination of hyTDG activity against defined substrates. However, a more robust assay would involve hyTDG activity against multiple substrates simultaneously. We therefore developed an approach that separates free bases and isotope-enriched internal standards from oligonucleotides or DNA and protein
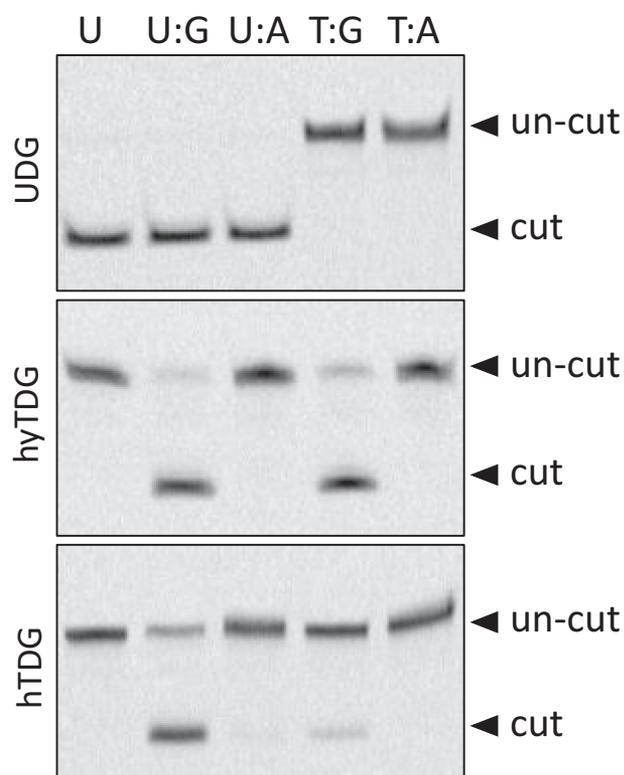
Figure 4. Comparison of UDG, hyTDG, and hTDG activity using a gel cleavage assay. Oligonucleotides with a 5′-FAM label (2.5 pmol) containing U or T at the target site were either unpaired or paired with complementary sequences to form U:A or T:A base pairs or U:G or T:G mispairs. Oligonucleotides were incubated for 1 h with UDG (2.5 units, 0.04 µg, 1.6 pmol) at 37 °C, hyTDG (0.5 µg, 16.8 pmol) at 65 °C, or hTDG (1.5 µg, 31.0 pmol) at 37 °C. Sodium hydroxide was then added to hydrolyze the phosphate backbone of oligonucleotides containing an abasic site, and oligonucleotides were resolved by PAGE, visualized, and quantified with a STORM imager. UDG cleaves all uracil-containing oligonucleotides completely but has no activity on T. The hyTDG cleaves only mispaired U (95%) and T (86%). The hTDG cleaves U mispaired with G (73%) but much less T mispaired with G (6%). 6-FAM, 6-carboxyfluorescein; hTDG, human thymine DNA glycosylase; hyTDG, hybrid thymine DNA glycosylase; UDG, uracil-DNA glycosylase.

Figure 5. Analysis of glycosylase activity of hyTDG on oligonucleotides using a real-time fluorescence assay. In the absence (*A*) and presence (*B*) of added calf thymus DNA. About 25 pmol duplexes with 5′-FAM (*upper strand*) and 3′-BHQ1 (*lower strand*) were incubated with 25 pmol hyTDG at 65 °C in the presence of DMDA, which cleaves abasic sites. Fluorescence was monitored in a Roche 480 qPCR instrument. The equation for the *solid lines* in each figure is $Y = A(1 - e^{-kt})$, where Y is the percent oligonucleotide cleaved, A is maximum percent cleaved, k is the rate constant (min^{-1}), and t is time in minutes. In panel *A*, the values of A and k for U:G are 96.2 ± 0.6% and 0.1075 ± 0.103 min^{-1} and for T:G are 92.4 ± 1.10% and 0.0694 ± 0.009 min^{-1}. In panel *B*, the experiment is identical to panel *A*, except that 20 µg of calf thymus DNA was added. Values of A and k in the presence of calf thymus for U:G are 99.9 ± 1.1% and 0.1136 ± 0.0147 min^{-1} and for T:G are 103 ± 1.1% and 0.0613 ± 0.0122 min^{-1}. The rate of U:G cleavage exceeds T:G, by factors of 1.6 to 1.9. The rate of U:G and T:G cleavage does not change significantly upon addition of calf thymus DNA. BHQ1, black hole fluorescence quencher 1; DMDA, *N,N*-dimethylethylenediamine; FAM, 6-carboxyfluorescein; hyTDG, hybrid thymine DNA glycosylase; qPCR, quantitative PCR.

using a spin filter. Isolated free bases can be chemically derivatized with t-butyldimethylsilyl (TBDMS) groups and analyzed by GC–MS/MS. This workflow is shown schematically in Figure 6.

We first applied this approach to a mixture of duplex oligonucleotides containing T:G and U:G mispairs in a two to one ratio. A mixture of 8.3 pmol U:G duplex, 16.7 pmol T:G duplex, and 250 pmol hyTDG with U + 3 and T + 4 standards in a volume of 25 µl was incubated at 65 °C for up to 120 min. The progress of the hyTDG reaction was followed simultaneously using both gel and GC–MS/MS methods (Fig. 7). A volume of 5 µl was used for the gel assay and 20 µl for the GC–MS/MS assay. Each time point was analyzed three times by GC–MS/MS.

As shown in Figure 7A, approximately 91% of the mispaired duplexes were cleaved in 120 min as measured by the gel assay. Base release was also monitored by GC–MS/MS analysis (Fig. 7B). Consistent with the gel assay, cleavage of both U and T appeared to plateau after 60 min. At 120 min, 6.42 ± 0.49 pmol of U was released and 13.33 ± 0.34 pmol of
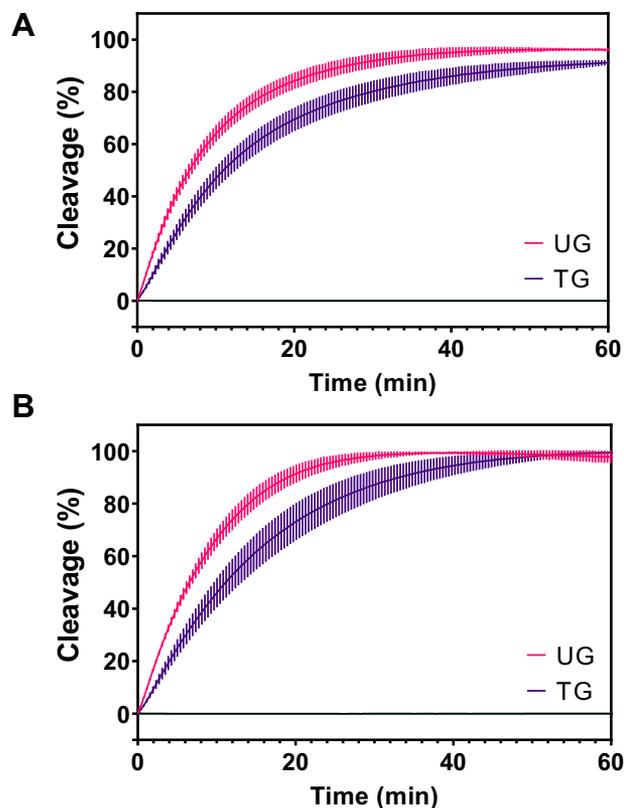
T was released. The amount of U and T released is consistent with the amount of U and T oligonucleotides in the reaction. As a control, the oligonucleotide mixture was also incubated with UDG at 37 °C for 120 min. Gel analysis indicated 43% cleavage, slightly higher than the 33% expected based upon the composition of the mixture. The amount of U released by UDG as measured by GC–MS/MS was 7.58 ± 0.20 pmol, also slightly higher than expected.

In a final series of experiments, the content of mispairs in calf thymus DNA was examined. First, calf thymus DNA was digested with the EcoRI restriction endonuclease to reduce its viscosity. Next, a portion of the calf thymus DNA was hydrolyzed in formic acid and the base composition examined by GC–MS using stable isotope–enriched standards of C, T, and 5-methylcytosine (5-mC). The base composition was observed to be 0.52 ± 0.04 nmol C, 0.78 ± 0.02 nmol T, and 0.03 ±
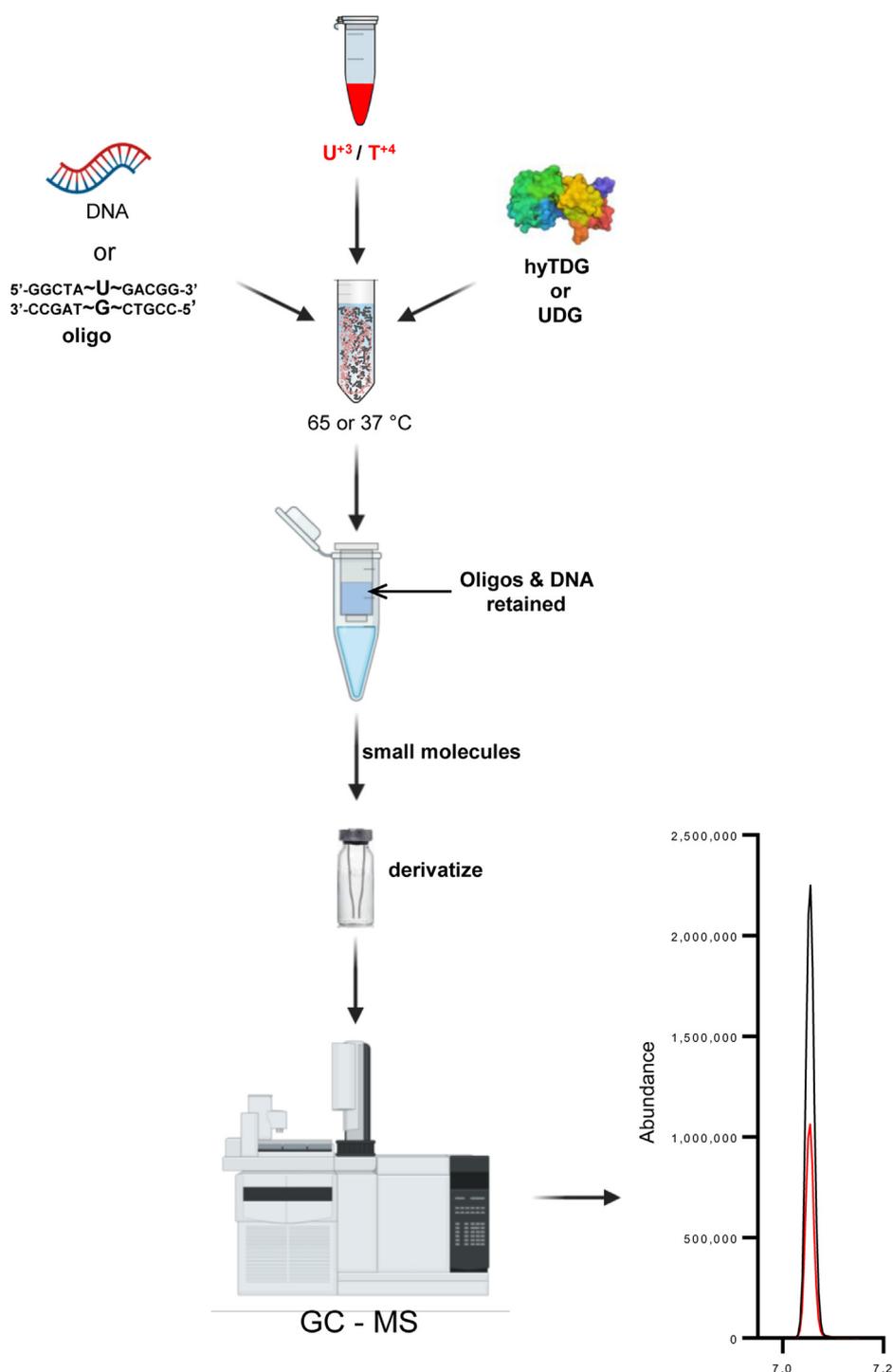
Figure 6. Workflow for measuring bases released by hyTDG or UDG using mass spectrometry. Oligonucleotides or DNA are incubated with hyTDG or UDG in the presence of one or more stable-isotope standards. Following incubation, free bases are isolated by spin filtration. Isolated bases are derivatized and analyzed by GC–EI–MS/MS or GC–NCI–MS. Pyrimidines released by the glycosylase are quantified by comparing the integrated peak area of the unenriched pyrimidine with the peak area of a corresponding stable isotope–enriched standard. Created in part with BioRender.com. EI, electron ionization; hyTDG, hybrid thymine DNA glycosylase; NCI, negative chemical ionization; UDG, uracil-DNA glycosylase.

0.0002 nmol 5-mC per microgram of calf thymus DNA, consistent with prior reports (35, 36).

To measure the content of U:G and T:G mispairs, a solution of EcoR1-digested calf thymus DNA (400 μg) containing isotope-enriched T + 4 (14.5 pg T + 4/μg DNA) and U + 3 (5 pg/μg DNA) was incubated with either UDG (37 °C) or hyTDG (65 °C) for 90 min. Released free bases were separated from DNA and enzymes by spin filtration. Filtrates were dried, and the pyrimidine composition was measured by two analytical approaches. In the first approach, pyrimidines released by the glycosylases were converted to the TBDMS derivatives and analyzed by GC–MS/MS. In the second approach, pyrimidines were converted to the 3,5-bis(trifluoromethyl)benzyl bromide derivatives and analyzed

SASBMB

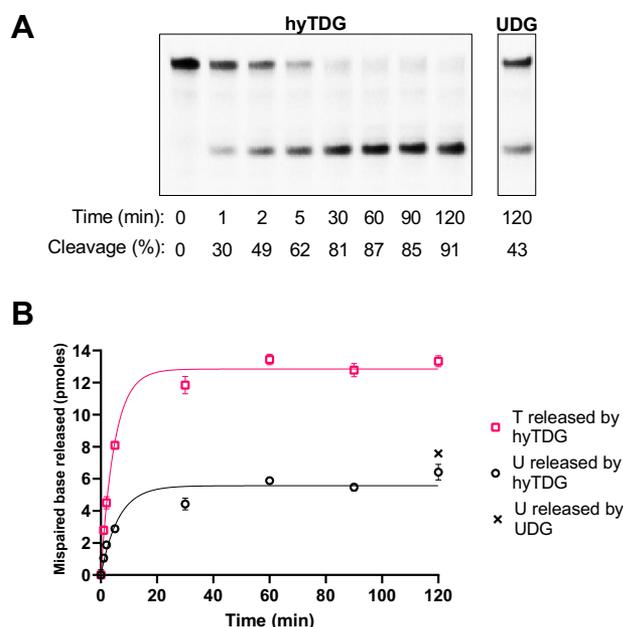*J. Biol. Chem.* (2022) 298(3) 101638 **5**

Figure 7. Cleavage of a mixture of oligonucleotides containing T:G and U:G mispairs by hyTDG followed simultaneously by gel electrophoresis and GC–MS/MS. Mixtures of 5′-FAM-labeled oligonucleotides containing U:G (8.3 pmol) or T:G (16.7 pmol) were incubated with 250 pmol hyTDG and isotope-enriched standards (U + 3, T + 4) in a total volume of 25 µl at 65 °C. At selected time intervals, 5 µl was used for gel electrophoresis, and the remaining 20 µl was used for the measurement of released bases. Released bases were separated by spin filtration, derivatized, and analyzed by GC–MS/MS. Gel analysis indicated predominant cleavage of U:G and T:G oligonucleotides by hyTDG by 60 min (panel *A*). The oligonucleotide mixture was also incubated with UDG at 37 °C (1 unit, 0.6 pmol), which cleaved the U:G but not T:G oligonucleotide (panel *A*, *far right*). Base release was measured by GC–MS/MS as shown in panel *B*. Each time point was analyzed three times. At 2 h, 6.42 ± 0.49 pmol U and 13.33 ± 0.34 pmol T were released, representing nearly complete release of U:G and T:G in the sample. The amount of U released by UDG at 2 h was 7.58 ± 0.2 pmol. FAM, 6-carboxyfluorescein; hyTDG, hybrid thymine DNA glycosylase.

Figure 8. Analysis of the release of U and mispaired T from calf thymus DNA by UDG and hyTDG. Approximately 400 µg of EcoRI-digested calf thymus DNA was incubated with UDG (10 units, 6.2 pmol, 37 °C) or hyTDG (295 pmol, 65 °C) for 90 min. Released bases were isolated by spin filtration, derivatized, and analyzed by GC–EI–MS/MS (panel *A*) or GC–NCI–MS (panel *B*). Data presented above represents observed amounts minus background from three independent experiments. In panel *A*, total uracil (single stranded, U:A and U:G) released by UDG was 9.39 ± 0.29 pg/µg DNA. The amount of uracil from U:G released by hyTDG was 1.30 ± 0.29 pg/µg, and the amount of T from T:G released was 5.58 ± 0.42 pg/µg. These amounts correspond to one deaminated U:G mispair per 4.48×10^4 C:G base pairs and one deaminated T:G mispair per 6.71×10^2 5-mC:G base pairs. In panel *B*, total uracil (single strand, U:A and U:G) released by UDG was 8.46 ± 0.63 pg/µg DNA. The amount of uracil from U:G released by hyTDG was 0.54 ± 0.13 pg/µg, and the amount of T from T:G released was 4.14 ± 0.21 pg/µg. These amounts correspond to one deaminated U:G mispair per 1.08×10^5 C:G base pairs and one deaminated T:G mispair per 9.09×10^2 5-mC:G base pairs. Most of the U is in U:A base pairs or single-stranded DNA (86% panel *A*, 94% panel *B*). The amount of T in T:G mispairs exceeds the amount of U in U:G mispairs by a factor of 4.3 (panel *A*) to 7.7 (panel *B*). EI, electron ionization; hyTDG, hybrid thymine DNA glycosylase; NCI, negative chemical ionization; UDG, uracil-DNA glycosylase.

by GC–MS using negative chemical ionization (GC–NCI–MS). All measurements for each approach represent three independent experiments.

Incubation with UDG releases uracil in U:A and U:G base pairs as well as in single-stranded DNA. Total uracil in the calf thymus DNA released by UDG was 9.39 ± 0.29 pg/µg DNA by GC–MS/MS (Fig. 8*A*). The amount of U released from U:G mispairs by hyTDG was 1.30 ± 0.29, and the amount of T:G released from T:G mispairs was 5.58 ± 0.42 pg/µg DNA. Chromatograms from the GC–electron ionization–MS/MS analyses are shown in Figure S12.

The amount of U and T released was also measured using GC–NCI–MS (Fig. 8, panel *B*). The amount of U released by UDG was measured to be 8.46 ± 0.63 pg/µg DNA. The amount of U released by hyTDG was measured to be 0.54 ± 0.13 pg/µg DNA, and the amount of T released was measured to be 4.14 ± 0.21 pg/µg DNA. Chromatograms for the GC–NCI–MS analyses are shown in Figure S13.

The experiments depicted in Figures 7 and 8 were conducted in the presence of U + 3 and T + 4 internal standards for GC–MS analysis. To ensure that the internal standards did not inhibit base excision by hyTDG, we conducted a gel-based assay under similar conditions except that up to 50 pmol of U
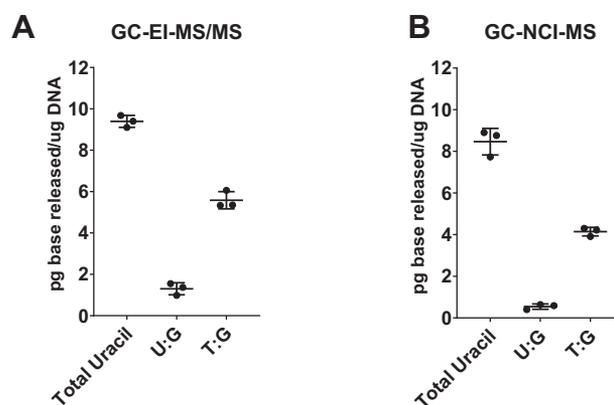
free base was added. The additional U free base had no observable effect upon glycosylase cleavage under the conditions of this experiment (Fig. S14).

Discussion

Cytosine to thymine transition mutations are abundant in human cancers and believed to result from the deamination of cytosine and its analogs (1–10). These mutations proceed through an intermediate mispair where the deaminated cytosine adduct or uracil analog is paired with guanine (Fig. 1). The deaminated mispair can either be repaired by excision repair pathways or converted to a mutation following DNA replication or repair synthesis. The relative rates of mispair formation, repair, and replication therefore determine the frequency with which the deamination of cytosine or its analogs results in mutation. Currently, adequate methods are lacking for measuring the deaminated intermediates.

To address this gap, we have constructed a hybrid glycosylase (hyTDG), which cleaves uracil and thymine, key deamination products, selectively from mispairs. The hybrid enzyme contains a 29 amino acid peptide from the hTDG linked to the catalytic domain of a thermophile TDG, MIG. The 29 amino acid N-terminal peptide of hTDG (residues 82–110) is unstructured and positively charged, which may promote nonspecific interactions with the DNA phosphate

backbone to promote lesion searching (25). Coey *et al.* (25) demonstrated that the addition of these residues to TDG[111–308], forming TDG[82–308], substantially increased the affinity and glycosylase activity to levels observed for the full-length hTDG. Thus, our rationale for linking the human peptide with a more active T:G glycosylase, MIG, was to enhance lesion scanning and facilitate overall glycosylase activity in the hybrid enzyme.

In contrast to hTDG, which cleaves U:G >> T:G, our hybrid enzyme has strong activity against both U:G and T:G mispairs, fulfilling the needed activity for our assay. Our hyTDG retains the mispair specificity previously reported for MIG (24). On oligonucleotide substrates, the hyTDG, which contains 29 amino acids from hTDG, is about 20% faster than MIG in the absence of calf thymus DNA. In the presence of excess calf thymus DNA, the ν_o for our hyTDG is approximately 2.3-fold higher than MIG. The primary advantage of the 29 amino acid peptide from hTDG appears to be in scanning genomic DNA to find its target.

The mispair selectivity of hyTDG, coupled with its high activity on T:G mispairs and ability to rapidly scan DNA, makes it a highly valuable tool for measuring mispaired U:G and T:G in genomic DNA. Using hyTDG and GC–MS quantitation methods, we have probed calf thymus DNA for the presence of U:G and T:G mismatches and UDG to measure total U.

Uracil can occur in DNA by two distinct mechanisms (37–40). The deamination of cytosine in a duplex would generate a U:G mispair. Alternatively, dUMP could be misincorporated by DNA polymerase into a U:A base pair during DNA replication. The amount of uracil in DNA from cytosine deamination (U:G) would increase with time and UDG deficiency. Uracil misincorporation can occur during DNA replication into U:A base pairs as polymerases show little discrimination against dUTP. Uracil in DNA from misincorporation of dUMP would increase from defects in one-carbon metabolism and deficiencies in UDG or dUTPase activity.

Previous methods to measure uracil in DNA have relied upon UDG release or DNA hydrolysis prior to analysis. Both methods measure total uracil. Importantly, the biological significance of uracil in DNA depends upon the base pairing context. Uracil in U:A base pairs reflects metabolic disturbances and if unrepaired could interfere with DNA–protein interactions (41–43), whereas uracil in a U:G mispair is promutagenic. Using two independent analytical approaches, we measured the total amount of uracil in calf thymus DNA to be approximately 9 pg/μg of DNA. We note that previous studies have reported levels as low as 0.2 pg/μg DNA in freshly isolated DNA (44) but can exceed 100 pg/μg in DNA isolated from cells with defects in one-carbon metabolism (45). Consistent with our findings, the level of uracil in salmon sperm DNA is reported to be approximately 4.5 pg/μg DNA (46).

Using the approach described here, we have determined the distribution of uracil between U:A and U:G base pairs in calf thymus DNA. We find that approximately 90% of the uracil in calf thymus DNA is in U:A base pairs, which result from dUMP misincorporation opposite A.

As with uracil, thymine could occur in a T:G base pair by deamination of 5-mC or by the misincorporation of T opposite G during DNA replication. In human cancer cells, C to T mutations occur with high frequency at CpG dinucleotides (47–49). In eukaryotic DNA, cytosine methylation occurs predominantly at CpG dinucleotides. In addition, most CpG dinucleotides are methylated in most tissues (50–52). While polymerase misincorporation could generate a T:G mispair, available data suggest that polymerase misincorporation or extension is not strongly sequence dependent (53, 54). The rate of deamination of 5-mC is slightly faster than cytosine (10, 55, 56). However, the repair of T:G mispairs in eukaryotic cells is slower than U:G mispairs by orders of magnitude (57). Therefore, the predominance of T:G mispairs in DNA likely arose from the deamination of 5-mC in CpG dinucleotides.

Using our method we have, for the first time, measured the level of T:G base pairs in DNA. We measured approximately 5 pg of T:G mispairs per microgram of DNA. The number of T:G mispairs exceeds that of U:G mispairs by a factor of 4 to 8. However, cytosine is approximately 17 times more abundant than 5-mC in calf thymus DNA. Therefore, a 5-mC:G base pair is 60 to 120 times more likely to be found as a T:G mispair than a C:G base pair found as a U:G mispair. Slow repair of T:G mispairs in eukaryotic cells (57) is consistent with their increased levels, despite 5-mC being 17-fold less abundant than C. The method reported here could allow measurement of the rates of formation and repair of T:G and U:G mispairs in cells.

Endogenous DNA damage, including deamination and oxidation, is an important source of mutation in human cells. It can also generate apparent "noise" in next-generation DNA-sequencing studies. Recently, several groups have sought to reduce damaged-related noise by incubating DNA with a cocktail of DNA repair enzymes prior to sequencing (58–64). A limitation of current approaches is that available repair enzymes do not efficiently act on the T:G mispair, which in our studies of calf thymus DNA is the most abundant mutagenic base pair examined here. The hyTDG enzyme described here should prove valuable in developing additional assays to both detect and study these frequent and critical DNA damage adducts.

Experimental procedures

Stable isotope standards

Enriched cytosine (C + 2, ²H₂, H5, and H6) and enriched 5-mC (5-mC + 4, methyl-²H₃, and H6) were obtained from CDN isotopes. Enriched thymine (T + 4, methyl ²H₃, and ²H6) was obtained from Cambridge Isotope Laboratories. Enriched uracil (U + 3, ¹⁵N₂, and ¹³C2) was obtained from Sigma–Aldrich.

Construction, cloning, and purification of the hyTDG

A DNA sequence was constructed with an amino-terminal His-tag (6xHis-tag), joined to the sequence encoding a 29

amino acid peptide from hTDG (amino acids 82–112, NM_003211.6) and the full-length TDG from *Methanothermobacter thermoautotrophicus* (MIG, *Orf 10*, WP_010889848.1). This hybrid DNA sequence was inserted into the pET-28a(+) expression vector between the NcoI and XhoI restriction sites. The hybrid DNA sequence is shown in Figure S1, and the corresponding amino acid sequence of the hyTDG is shown in Figure 2.

The pET-28a(+)-hyTDG plasmid was transformed into *Escherichia coli* strain BL21 (DE3). Transformants were selected on an agar plate containing kanamycin. Selected clones were grown in 100 ml LB broth supplemented with kanamycin and induced with IPTG for 6 h at 30 °C. Cells were harvested by centrifugation at 4100 rpm for 5 min and stored at −20 °C until used. Cell pellets were thawed and suspended in 4 ml lysis buffer (50 mM potassium phosphate, 20 mM imidazole, 3000 mM sodium chloride, 10 mM β-mercaptoethanol, 1% Triton, and 1 mM PMSF) and sonicated for eight cycles, 30 s each with 30 s breaks on ice.

Supernatants were then centrifuged (12,000 rpm, 10 min), loaded onto previously equilibrated nickel-charged resin (HisPur Ni–NTA resin; catalog no.: 88221; Thermo Fisher Scientific), and incubated for 1.5 h at 4 °C. The resin and supernatant were centrifuged on a column at 1000*g* and washed as recommended by the vendor. The bound His-tagged protein was eluted with buffer (50 mM potassium phosphate, 300 mM sodium chloride, 10 mM β-mercaptoethanol, and 100 mM imidazole). Total protein concentration was measured with a Bradford protein bioassay. Isolated protein was analyzed on a 12% Tris-glycine polyacrylamide gel stained with Coomassie blue (Fig. S2), which indicated an apparent molecular weight of 26.5 kDa.

Characterization of the purified hyTDG by LC–MS/MS analysis

Approximately 10 μg of recombinant hyTDG was purified by SDS-PAGE. Gel bands were cut from the gel, destained with 50% methanol in water, and dried. Gel bands were resuspended in 50 μl acetic anhydride and 200 μl acetic acid to chemically acetylate protein lysine residues. After incubation at 37 °C for 1 h, the liquid was removed, and gel bands were washed three times with 1 ml deionized water. Gel bands were then dried and ground into a fine powder. Ammonium bicarbonate solution (100 μl, 50 mM) was added, and the pH of the resulting gel was increased to approximately 8 with aqueous ammonia. Trypsin was then added, and the proteins were digested overnight at 37 °C.

Tryptic peptides were extracted with acetonitrile, dried, and resuspended in 50 μl of 1% formic acid for LC–MS/MS analysis. Tryptic peptides were loaded onto a reversed-phase ProteoPre column loaded with Waters 5 μ XSelect HSS T3 resin and Waters YMC ODS-AQ S-5 100 A resin and eluted with a gradient of acetonitrile in 0.1% formic acid. The LC column was directly interfaced with a QExactive mass analyzer, which acquired data at a resolution of 35,000 in full scan mode and 17,500 in MS/MS mode. The top most intense peptides in each MS survey were selected for MS/MS analysis.

Peptides were identified with the PEAKS 8.5 software for *de novo* peptide sequencing. A maximum of three missed cleavages was set. Acetylation of K, S T, C, Y and H as well as oxidation of methionine and deamination of asparagine and glutamine were set as variable modifications. Otherwise, no fixed modifications were set. Parent mass tolerance was set to 20.0 ppm, and fragment mass was set to 0.05 Da.

Gel-based cleavage assay

A series of oligonucleotides (Fig. S8) were constructed containing a central pyrimidine, X, cytosine (C), uracil (U), or thymine (T) paired opposite a purine (P), adenine (A), or guanine (G). One sequence (5′-6FAM-CGT-GGC-XGG-CCA-CGA-CGG-3′) contained the fluorophore, 6FAM on the 5′ end. The complementary strand (5′-CCG-TCG-TGG-CCP-GCC-ACG) was synthesized with and without the 3′-BHQ1 synthesized with 4′-(2-nitro-4-toluyldiazo)-2′-methoxy-5′-methyl-azobenzene-4″-(N-ethyl-2-O-(4,4′-dimethoxytrityl))-N-ethyl-2-O-glycolate-linked controlled pore glass resin.

In a typical assay examined by gel electrophoresis, 2.5 pmol of 5′-6FAM-labeled oligonucleotide and two equivalents of an unlabeled complementary sequence were incubated in 10 μl buffer (10 mM potassium phosphate, 30 mM sodium chloride, and 40 mM potassium chloride) with UDG (2.5 units, 0.04 μg, 1.6 pmol, *E. coli* UDG, New England Biolabs, 37 °C), hyTDG (0.5 μg, 16.8 pmol, 65 °C), or hTDG (1.5 μg, 31.0 pmol, 37 °C,) for 1 h (30, 31). The reaction was terminated, and the phosphate backbone of the oligonucleotide containing an abasic site was cleaved with 2 μl and 1 M NaOH at 95 °C for 10 min. Formamide (10 μl) was then added, and the reaction mixture was loaded onto a 6 M urea denaturing 20% polyacrylamide gel. The oligonucleotide mixture was resolved by electrophoresis for 45 min. Gels containing fluorescent bands were visualized and quantified on a Storm 860 phosphorimager.

Real-time fluorescence assay

In a typical real-time florescence assay, 25 pmol of 5′-6FAM-labeled oligonucleotide was annealed with 50 pmol of the complementary sequence containing the 3′-BHQ1 quencher in a 25 μl reaction volume containing 10 mM potassium phosphate buffer, pH 7.7, 30 mM NaCl, and 40 mM KCl. To ensure cleavage of the phosphate backbone following glycosylase release of a target base, DMDA (final concentration of 100 mM) was added. The reaction was initiated upon the addition of the glycosylase, and fluorescence was monitored at 65 °C every 20 s in a Roche 480 quantitative PCR instrument. Real-time fluorescence assays were acquired in triplicate. Graphs of data were prepared with the PRISM software (GraphPad).

Oligonucleotide cleavage assays monitored by gel and GC–MS/MS

To follow cleavage by gel electrophoresis and GC–MS/MS, a series of mixture containing 8.3 pmol 5′-6FAM-labeled U:G oligonucleotide, 16.7 pmol 5′-6FAM-labeled T:G oligonucleotide, U + 3 (8.3 pmol) and T + 4 (16.7 pmol) internal

standards, and 250 pmol hyTDG in buffer were incubated at 65 °C in a volume of 25 μl.

At selected time intervals, 5 μl of the reaction mixture was taken for gel electrophoresis and 20 ml for GC–MS/MS analysis. Oligonucleotides were analyzed by gel electrophoresis as described previously. Samples examined by GC–MS/MS were first diluted to 400 μl with water and spin-filtered (catalog no.: UFC500396; Amicon Ultra Ultracel 3k) at 14,000g for 45 min. Solutions were then dried under reduced pressure.

Pyrimidines released by hyTDG were converted to their TBDMS derivatives in acetonitrile, and 0.5 μl of the reaction solution was injected onto an Agilent 7890B GC containing an HP-5 column. The GC oven temperature was held constant at 100 °C for 2 min, ramped to 260 °C at 30 °C/min, and held at that temperature for 10 min. The GC was directly coupled to an Agilent 7000C triple quadrupole detector. The most predominant ions of both uracil (283 amu) and thymine (297 amu) derivatives correspond to the M-57 (*t*-butyl) fragment. The transitions used to monitor uracil were 283 to 169 amu and for thymine 297 to 183. The corresponding transitions for the U + 3 internal standard were 3 amu higher than U, and for T + 4, 4 amu higher than T.

Preparation of calf thymus DNA and analysis of base composition

Calf thymus DNA was dissolved in buffer containing 5 mM NaCl, 1 mM Tris (pH 7), 1 mM $MgCl_2$, and 0.1 mM DDT. DNA (∼50 mg) was digested with ∼20,000 units of EcoRI endonuclease (New England Biolabs) at 37 °C for 4 h to reduce viscosity ([64]). Digested DNA was precipitated with ammonium acetate/ethanol, resuspended in buffer, and dialyzed overnight.

A portion of the digested calf thymus DNA was hydrolyzed in 88% formic acid at 140 °C for 40 min. Isotope-enriched standards of thymine (T + 4), cytosine (C + 2), and 5-mC (5-mC + 3) at a ratio of 20:1 (C/5-mC) were added to the vials, which were then evaporated to dryness under reduced pressure. Bases were converted to the TBDMS derivatives in acetonitrile at 140 °C for 40 min. Samples were injected onto an Agilent 7890A GC containing a DB5 column. The initial GC oven temperature was 100 °C for 2 min, ramped to 260 °C at 30 °C per min, and then held at 260 °C for 10 min. The GC was directly interfaced to an Agilent 5975C mass selective detector, and data were collected in the selected ion mode. Molar amounts of C and T were determined by comparing experimental peak areas to standard curves. The molar amount of 5-mC was determined by comparing peak areas of unenriched C and 5-mC to peak areas of the isotope-enriched standards. Base composition determinations were done in triplicate.

Analysis of bases released from calf thymus DNA by UDG and hyTDG

EcoRI-digested DNA was dissolved in buffer containing 10 mM phosphate (pH 7.7), 30 mM NaCl, and 40 mM KCl, and isotope-enriched internal standards were added (14.5 pg T + 4/μg DNA and 5 pg U + 3/μg DNA). Approximately 400 μg of DNA plus internal standards was then incubated with UDG (10 units, 6.2 pmol, 37 °C, 2 h) or hyTDG (295 pmol, 65 °C, 2 h). Reaction mixtures were spin filtered, and the filtrate was dried under reduced pressure.

Two analytical approaches were used to measure released pyrimidines. In the first approach, released pyrimidines were converted to the TBDMS derivatives and analyzed by GC–MS/ MS as described previously.

In the second approach, released pyrimidines were converted to the 3,5-bis(trifluoromethyl)benzyl bromide derivatives and analyzed by GC–MS using NCI (GC–NCI–MS). Data were collected in the selected ion mode for the ions 337 *m/z* (U), 340 *m/z* (U + 3), 351 *m/z* (T), and 355 *m/z* (T + 4). The amounts of uracil and thymine in each sample were determined by comparing the integrated peak areas for each with the integrated peak areas of the corresponding isotope-enriched internal standards. Levels of released T and U were then obtained by subtracting the background levels of the pyrimidines in control calf thymus DNA and glycosylase solutions. All measurements were made in triplicate

Data availability

All data are contained within the article and supporting information. The MS proteomics data have been deposited to the ProteomeXchange Consortium *via* the PRIDE partner repository ([65]) with the dataset identifier PXD030459 and 10.6019/PXD030459.

Enzymatic repair of deaminated cytosine adducts

References

1. Hollstein, M., Sidransky, D., Vogelstein, B., and Curtis, H. C. (1991) p53 mutations in human cancers. *Science* **253**, 49–53
2. Magewu, A. N., and Jones, P. A. (1994) Ubiquitous and tenacious methylation of the CpG site in codon 248 of the p53 gene may explain its frequent appearance as a mutational hot spot in human cancer. *Mol. Cell. Biol.* **14**, 4225–4232
3. Iengar, P. (2012) An analysis of substitution, deletion and insertion mutations in cancer genes. *Nucleic Acids Res.* **40**, 6401–6413
4. Forbes, S. A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C. G., Ward, S., Dawson, E., Ponting, L., Stefancsik, R., Harsha, B., YinKok, C., Jia, M., Jubb, H., *et al.* (2017) Cosmic: Somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45**, D777–D783
5. Lewis, C. A., Crayle, J., Zhou, S., Swanstrom, R., and Wolfenden, R. (2016) Cytosine deamination and the precipitous decline of spontaneous mutation during Earth's history. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 8194–8199
6. Lindahl, T., and Nyberg, B. (1974) Heat-induced deamination of cytosine residues in DNA. *Biochemistry* **13**, 3405–3410
7. Coulondre, C., Miller, J. H., Farabaugh, P. J., and Gilbert, W. (1978) Molecular basis of base substitution hotspots in Escherichia coli. *Nature* **274**, 775–780
8. Duncan, B. K., and Miller, J. H. (1980) Mutagenic deamination of cytosine residues in DNA. *Nature* **287**, 560–561
9. Wang, R. Y. H., Kuo, K. C., Gehrke, C. W., Huang, L.-H., and Ehrlich, M. (1982) Heat and Alkali-induced deamination of 5-methylcytosine and cytosine residues in DNA. *Biochim. Biophys. Acta* **697**, 371–377
10. Shen, J. C., Rideout, W. M., and Jones, P. A. (1994) The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res.* **22**, 972–976
11. Cadet, J., and Wagner, J. R. (2013) DNA base damage by reactive oxygen species, oxidizing agents, and UV radiation. *Cold Spring Harb. Perspect. Biol.* **5**, 1–18
12. Sangaraju, D., Villalta, P. W., Wickramaratne, S., Swenberg, J., and Tretyakova, N. (2014) NanoLC/ESI+ HRMS3 quantitation of DNA adducts induced by 1,3-butadiene. *J. Am. Soc. Mass Spectrom.* **25**, 1124–1135
13. You, C., and Wang, Y. (2016) Mass spectrometry-based quantitative strategies for assessing the biological consequences and repair of DNA adducts. *Acc. Chem. Res.* **49**, 205–213
14. Jumpathong, W., Chan, W., Taghizadeh, K., Babu, I. R., and Dedon, P. C. (2015) Metabolic fate of endogenous molecular damage: Urinary glutathione conjugates of DNA-derived base propenals as markers of inflammation. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E4845–E4853
15. Gates, K. S. (2009) An overview of chemical processes that damage cellular DNA: Spontaneous hydrolysis, alkylation, and reactions with radicals. *Chem. Res. Toxicol.* **22**, 1747–1760
16. Totsuka, Y., Watanabe, M., and Lin, Y. (2021) New horizons of DNA adductome for exploring environmental causes of cancer. *Cancer Sci.* **112**, 7–15
17. Blount, B. C., and Ames, B. N. (1994) Analysis of uracil in DNA by gas chromatography-mass spectrometry. *Anal. Biochem.* **219**, 195–200
18. Beckman, K. B., Saljoughi, S., Mashiyama, S. T., and Ames, B. N. (2000) A simpler, more robust method for the analysis of 8-oxoguanine in DNA. *Free Radic. Biol. Med.* **29**, 357–367
19. Jaruga, P., Kirkali, G., and Dizdaroglu, M. (2008) Measurement of formamidopyrimidines in DNA. *Free Radic. Biol. Med.* **45**, 1601–1609
20. Mullins, E. A., Rubinson, E. H., Pereira, K. N., Calcutt, M. W., Christov, P. P., and Eichman, B. F. (2013) An HPLC-tandem mass spectrometry method for simultaneous detection of alkylated base excision repair products. *Methods* **64**, 59–66
21. Minko, I. G., Vartanian, V. L., Tozaki, N. N., Coskun, E., Coskun, S. H., Jaruga, P., Yeo, J., David, S. S., Stone, M. P., Egli, M., Dizdaroglu, M., McCullough, A. K., and Lloyd, R. S. (2020) Recognition of DNA adducts by edited and unedited forms of DNA glycosylase NEIL1. *DNA Repair (Amst)* **85**, 102741
22. Waters, T. R., and Swann, P. F. (1998) Kinetics of the action of thymine DNA glycosylase. *J. Biol. Chem.* **273**, 20007–20014
23. Bennett, M. T., Rodgers, M. T., Hebert, A. S., Ruslander, L. E., Eisele, L., and Drohat, A. C. (2006) Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. *J. Am. Chem. Soc.* **128**, 12510–12519
24. Liu, P., Burdzy, A., and Sowers, L. C. (2002) Substrate recognition by a family of uracil-DNA glycosylases: UNG, MUG, and TDG. *Chem. Res. Toxicol.* **15**, 1001–1009
25. Coey, C. T., Malik, S. S., Pidugu, L. S., Varney, K. M., Pozharski, E., and Drohat, A. C. (2016) Structural basis of damage recognition by thymine DNA glycosylase: Key roles for N-terminal residues. *Nucleic Acids Res.* **44**, 10248–10258
26. Horst, J. P., and Fritz, H. J. (1996) Counteracting the mutagenic effect of hydrolytic deamination of DNA 5-methylcytosine residues at high temperature: DNA mismatch N-glycosylase Mig.Mth of the thermophilic archaeon Methanobacterium thermoautotrophicum THF. *EMBO J.* **15**, 5459–5469
27. Begley, T. J., and Cunningham, R. P. (2002) Methanobacterium thermoformicicum thymine DNA mismatch glycosylase: Conversion of an N-glycosylase to an AP lyase. *Protein Eng. Des. Sel.* **12**, 333–340
28. Mol, C. D., Arvai, A. S., Begley, T. J., Cunningham, R. P., and Tainer, J. A. (2002) Structure and activity of a thermostable thymine-DNA glycosylase: Evidence for base twisting to remove mismatched normal DNA bases. *J. Mol. Biol.* **315**, 373–384
29. Yoon, J. H., Iwai, S., O'Connor, T. R., and Pfeifer, G. P. (2003) Human thymine DNA glycosylase (TDG) and methyl-CpG-binding protein 4 (MBD4) excise thymine glycol (Tg) from a Tg:G mispair. *Nucleic Acids Res.* **31**, 5399–5404
30. Hardeland, U., Bentele, M., Jiricny, J., and Schar, P. (2000) Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *J. Biol. Chem.* **275**, 33449–33456
31. Schuermann, D., Scheidegger, S. P., Weber, A. R., Bjørås, M., Leumann, C. J., and Schär, P. (2016) 3CAPS -A structural AP-site analogue as a tool to investigate DNA base excision repair. *Nucleic Acids Res.* **44**, 2187–2198
32. Kladova, O. A., Iakovlev, D. A., Groisman, R., Ishchenko, A. A., Saparbaev, M. K., Fedorova, O. S., and Kuznetsov, N. A. (2020) An assay for the activity of base excision repair enzymes in cellular extracts using fluorescent DNA probes. *Biochemistry* **85**, 480–489
33. Mechetin, G. V., Endutkin, A. V., Diatlova, E. A., and Zharkov, D. O. (2020) Inhibitors of DNA glycosylases as prospective drugs. *Int. J. Mol. Sci.* **21**, 1–30
34. Mchugh, P. J., and Knowland, J. (1995) Novel reagents for chemical cleavage at abasic sites and UV photoproducts in DNA. *Nucleic Acids Res.* **23**, 1664–1670
35. Kirk, J. T. (1967) Determination of the base composition of deoxyribonucleic acid by measurement of the adenine-granine ratio. *Biochem. J.* **105**, 673–677
36. Sturm, K. S., and Taylor, J. H. (1981) Distribution of 5-methylcytosine in the DNA of somatic and germline cells from bovine tissues. *Nucleic Acids Res.* **9**, 4537–4546
37. Richards, R. G., Sowers, L. C., Laszlo, J., and Sedwick, W. D. (1984) The occurrence and consequences of deoxyuridine in DNA. *Adv. Enzyme Regul.* **22**, 157–185
38. Kavli, B., Otterlei, M., Slupphaug, G., and Krokan, H. E. (2007) Uracil in DNA-general mutagen, but normal intermediate in acquired immunity. *DNA Repair (Amst)* **6**, 505–516
39. Olinski, R., Jurgowiak, M., and Zaremba, T. (2010) Uracil in DNA-Its biological significance. *Mutat. Res.* **705**, 239–245

ASBMB

40. Dube, D. K., Kunkel, T. A., Seal, G., and Loeb, L. A. (1979) Distinctive properties of mammalian DNA polymerases. *Biochim. Biophys. Acta* **561**, 369–382

41. Ivarie, R. (1987) Thymine methyls and DNA-protein interactions. *Nucleic Acids Res.* **15**, 9975–9983

42. Pu, W. T., and Struhl, K. (1992) Uracil interference, a rapid and general method for defining protein-DNA interactions involving the 5-methyl group of thymines: The GCN4-DNA complex. *Nucleic Acids Res.* **20**, 771–775

43. Rogstad, D. K., Liu, P., Burdzy, A., Lin, S. S., and Sowers, L. C. (2002) Endogenous DNA lesions can inhibit the binding of the AP-1 (c-Jun) transcription factor. *Biochemistry* **41**, 8093–8102

44. Ren, J., Ulvik, A., Refsum, H., and Ueland, P. M. (2002) Uracil in human DNA from subjects with normal and impaired folate status as determined by high-performance liquid chromatography-tandem mass spectrometry. *Anal. Chem.* **74**, 295–299

45. Mashiyama, S. T., Hansen, C. M., Roitman, E., Sarmiento, S., Leklem, J. E., Shultz, T. D., and Ames, B. N. (2008) An assay for uracil in human DNA at baseline: Effect of marginal vitamin B6 deficiency. *Anal. Biochem.* **372**, 21–31

46. Galashevskaya, A., Sarno, A., Vågbø, C. B., Aas, P. A., Hagen, L., Slupphaug, G., and Krokan, H. E. (2013) A robust, sensitive assay for genomic uracil determination by LC/MS/MS reveals lower levels than previously reported. *DNA Repair (Amst)* **12**, 699–706

47. Mancini, D., Singh, S., Ainsworth, P., and Rodenhiser, D. (1997) Constitutively methylated CpG dinucleotides as mutation hot spots in the retinoblastoma gene (RB1). *Am. J. Hum. Genet.* **61**, 80–87

48. Cooper, D. N., Mort, M., Stenson, P. D., Ball, E. V., and Chuzhanova, N. A. (2010) Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in CpNpG trinucleotides, as well as in CpG dinucleotides. *Hum. Genomics.* **4**, 406–410

49. Poulos, R. C., Olivier, J., and Wong, J. W. H. (2017) The interaction between cytosine methylation and processes of DNA replication and repair shape the mutational landscape of cancer genomes. *Nucleic Acids Res.* **45**, 7786–7795

50. Tornaletti, S., and Pfeifer, G. P. (1995) Complete and tissue-independent methylation of CpG sites in the p53 gene: Implications for mutations in human cancers. *Oncogene* **10**, 1493–1499

51. Rideout, W. M., Coetzee, G. A., Olumi, A. F., Spruck, C. H., and Jones, P. A. (1991) 5-Methylcytosine as an endogenous mutagen in the p53 tumor suppressor gene. *Science* **249**, 1288–1290

52. Jang, H. S., Shin, W. J., Lee, J. E., and Do, J. T. (2017) CpG and non-CpG methylation in epigenetic gene regulation and brain function. *Genes (Basel).* **8**, 2–20

53. Dosanjh, M. K., Galeros, G., Singer, B., and Goodman, M. F. (1991) Kinetics of extension of o6-methylguanine paired with cytosine or thymine in defined oligonucleotide sequences. *Biochemistry* **30**, 11595–11599

54. Shen, J. C., Creighton, S., Jones, P. A., and Goodman, M. F. (1992) A comparison of the fidelity of copying 5-methylcytosine and cytosine at a difined DNA template site. *Nucleic Acids Res.* **20**, 5119–5125

55. Sowers, L. C., David Sedwick, W., and Shaw, B. R. (1989) Hydrolysis of N3-methyl-2′-deoxycitidine: Model compound for reactivity of protonated cytosine residues in DNA. *Mutat. Res.* **215**, 131–138

56. Ehrlich, M., Norris, K. F., Wang, R. Y., Kuo, K. C., and Gehrke, C. W. (1986) DNA cytosine methylation and heat-induced deamination. *Biosci. Rep.* **6**, 387–393

57. Schmutte, C., Yang, A. S., Beart, R. W., and Jones, P. A. (1995) Base excision repair of U:G mismatches at a mutational Hotspot in the p53 gene is more efficient than base excision repair of T:G mismatches in extracts of human colon tumors. *Cancer Res.* **55**, 3742–3746

58. Briggs, A. W., and Heyn, P. (2012) Preparation of next-generation sequencing libraries from damaged DNA. *Methods Mol Biol* **840**, 143–154

59. Do, H., Wong, S. Q., Li, J., and Dobrovic, A. (2013) Reducing sequence artifacts in amplicon-based massively parallel sequencing of formalin-fixed paraffin-embedded DNA by enzymatic depletion of uracil-containing templates. *Clin. Chem.* **59**, 1376–1383

60. Costello, M., Pugh, T. J., Fennell, T. J., Stewart, C., Lichtenstein, L., Meldrim, J. C., Fostel, J. L., Friedrich, D. C., Perrin, D., Dionne, D., Kim, S., Gabriel, S. B., Lander, E. S., Fisher, S., and Getz, G. (2013) Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res.* **41**, 1–12

61. Arbeithuber, B., Makova, K. D., and Tiemann-Boege, I. (2016) Artifactual mutations resulting from DNA lesions limit detection levels in ultra-sensitive sequencing applications. *DNA Res.* **23**, 547–559

62. Kim, S., Park, C., Ji, Y., Kim, D. G., Bae, H., van Vrancken, M., Kim, D. H., and Kim, K. M. (2017) Deamination effects in formalin-fixed, paraffin-embedded tissue samples in the era of precision medicine. *J. Mol. Diagn.* **19**, 137–146

63. Chen, L., Liu, P., Evans, T. C., Jr., and Ettwiller, L. M. (2017) DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification. *Science* **355**, 752–756

64. Philippsen, P., Streeck, R. E., and Zachau, H. G. (1975) Investigation of the repetitive sequences in calf DNA by cleavage with restriction nucleases. *Eur. J. Biochem.* **57**, 55–68

65. Vizcaíno, J. A., Côté, R. G., Csordas, A., Dianes, J. A., Fabregat, A., Foster, J. M., Griss, J., Alpi, E., Birim, M., Contell, J., O'Kelly, G., Schoenegger, A., Ovelleiro, D., Pérez-Riverol, Y., Reisinger, F., *et al.* (2013) The Proteomics Identifications (PRIDE) database and associated tools: Status in 2013. *Nucleic Acids Res.* **41**, D1063–D1069