Base Analog and Neighboring Base Effects on Substrate Specificity of Recombinant Human G:T Mismatch-Specific Thymine DNA−Glycosylase†

Sibghat-Ullah,‡ Paola Gallinari,§ Yao-Zhong Xu,|| Myron F. Goodman,⊥ Linda B. Bloom,⊥ Josef Jiricny,§ and Rufus S. Day, III*,‡

Molecular Oncology Program, Cross Cancer Institute, Edmonton, Alberta, Canada T6G 1Z2, Cancer Research Campaign Nitrosamine-Induced Cancer Research Group, Department of Biochemistry and Molecular Biology, University College London, Grower Street, London WC1E 6BT, England, Department of Biological Sciences, Molecular Section, University of Southern California, Los Angeles, California 90089-1340, and Istituto di Ricerche di Biologia Molecolare, P. Angeletti, Via Pontina km. 30.600, 00040-Pomezia (Roma), Italy

Received April 30, 1996; Revised Manuscript Received July 16, 1996⊗

ABSTRACT: We studied the substrate specificity of the human G:T mismatch-specific thymine glycosylase that initiates the repair of G:T and G:U base mismatches to G:C base pairs. Such mismatches arise when 5-methylcytosine or cytosine deaminate spontaneously (and hydrolytically) in DNA. Substrates were 45-bp DNA heteroduplexes that bore single G:T, m6G:T, 2,6-diaminopurine:T, 2-amino-6-(methylamino)-purine:T, 2-aminopurine:T, and G:m4T mispairs. The bases 5′ to the poorly matched G were altered in selected G:T substrates to yield mispairs in four different contexts, ApG, CpG, GpG, and TpG. The recombinant thymine glycosylase was incubated with the 45-bp DNA substrates, each labeled at the 5′-terminus of the strand containing the mismatched T. The DNAs were then treated with 0.1 N NaOH to catalyze phosphodiester bond breakage at the newly-generated AP sites, and the products were analyzed on DNA sequencing gels. As indicated by the amounts of the 20-nt incision product, the removal of the thymine base by the enzyme increased linearly between 0 and 40 min at which time the generation of product from all substrates ceased, probably because of enzyme inactivation. The rate of incision was greatest (0.7 fmol/min) with DNA containing the G:T mispair followed by the DNA containing the m6G:T mispair (0.38 fmol/min) and the DNA with the 2-amino-6-(methylamino)purine:T mispair (0.15 fmol/min); the extent of reaction was 90%, 40%, and 20% respectively. By contrast to previous findings with cell-free extracts, DNA substrates containing 2,6-diaminopurine:T, 2-aminopurine:T, and G:m4T mispairs were not incised (<2%). The amount of incision of the 45-bp DNA substrates containing G:T mispairs in the CpG context was 3−12-fold greater than in the TpG, GpG, and ApG contexts.

DNA base mismatches are produced by errors of replication, as intermediates during recombination and by spontaneous deamination of m5C[1] in DNA (Duncan & Miller, 1980; Claverys & Lacks, 1986; Radman, 1988; Frederico et al., 1990; Modrich, 1994). G:T and the seven other possible mispairs that are produced as replication errors are repaired with different efficiencies by extracts of human cells (Holmes et al., 1990; Thomas et al., 1991) *via* a mismatch repair process analogous to the long-patch Mut HLS system of *Escherichia coli* (Modrich, 1991). By contrast, spontaneous hydrolytic deamination of m5C and cytosine in double-stranded DNA (Shapiro & Klein, 1966; Frederico et al., 1990) produce exclusively G:T and G:U mispairs respectively and independently of semiconservative replication or genetic recombination. In *E. coli*, G:T mispairs arising in this manner are repaired by a mismatch-specific endonuclease, Vsr (Hennecke et al., 1991), whereas in human cells, they are addressed by a mismatch-specific thymine/uracil DNA glycosylase (Wiebauer & Jiricny, 1989, 1990; Neddermann & Jiricny, 1993). In vertebrates, CpG dinucleotides are sites of deoxycytidine methylation (Razin & Riggs, 1980), and thus G:T mispairs arising from spontaneous hydrolytic deamination of m5C occur solely at CpG sites; G:T mismatches in other sequence contexts must have arisen as errors of replication or during recombination and therefore might not be addressed by the G:T glycosylase *in vivo*.

After studies in human cells demonstrated efficient repair of G:T mispairs to G:C pairs (Brown & Jiricny, 1987), *in vitro* studies identified specific repair of G:T to G:C pairs by a short-patch repair mechanism (Wiebauer & Jiricny, 1989, 1990). Human short-patch repair is initiated by a mismatch-specific thymine/uracil DNA glycosylase (Neddermann & Jiricny, 1993) that specifically removes the mismatched T (or U) from the DNA (Neddermann & Jiricny, 1994). An AP endonuclease then incises the DNA at the apyrimidinic site (Levin & Demple, 1990) and the baseless sugar-phosphate is removed by an as yet uncharacterized enzyme such that a one nucleotide gap appears in the DNA at the site of the mismatch (Wiebauer & Jiricny, 1989). This is filled in by DNA polymerase β (Wiebauer & Jiricny, 1990) and the nick is finally sealed by ligation. The mismatch-specific thymine DNA glycosylase has been characterized as a 55 kDa protein lacking AP-endonuclease activity that

† This work was supported by NIH Grants CA49936 to R.S.D. and GM21422 and AG11398 to M.F.G.

* To whom correspondence should be addressed. Tel: (403) 432-8898. FAX: (403) 432-8892.

‡ Cross Cancer Institute.

§ Istituto di Recerche di Biologia Molecolare.

|| University College London.

⊥ University of Southern California.

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

[1] Abbreviations: m5C, 5-methylcytosine; m6G, O^6-methylguanine; m4T, O^4-methylthymine; DiAP, 2,6-diaminopurine; AMAP, 2-amino-6-(methylamino)purine; 2AP, 2-aminopurine; AP, abasic (apurinic, apyrimidinic); MGMT, O^6-methylguanine−DNA methyltransferase.

S0006-2960(96)01022-7 CCC: $12.00

acts on G:U > G:T > > > C:T > T:T mispairs (Neddermann & Jiricny, 1993, 1994).

Several laboratories are seeking to understand the effects on the short-patch repair activity of the DNA sequence surrounding the G:T mispair. Studies of one DNA sequence (Griffin & Karran, 1993) showed that extracts of human cells repaired G:T mispairs to G:C pairs only when the G:T was present as 5′-CpG-3′/3′-GpT-5′. The G:T mispairs 5′-ApG-3′/3′TpT-5′, 5′-GpG-3′/3′-CpT-5′, and 5′-TpG-3′/3′-ApT-5′ were incised much less efficiently. More recent studies (Sibghat-Ullah & Day, 1995a) have shown that the repair of a G:T at one DNA site occurs in a CpG sequence-specific manner as Griffin and Karran (1993) had determined, but that at another site G:T mismatches were repaired independently of the base 5′ to the mismatched G, in agreement with an *in vivo* study (Brown & Jiricny, 1987). This may imply that repair in the short-patch mode acts both on G:T pairs produced by spontaneous deamination of m5C and on selected G:T mispairs in DNA produced as DNA biosynthetic errors or recombination intermediates.

In previous studies, A1235 human malignant glioma cell line cell-free extracts incised short double-stranded DNAs containing m6G:T mispairs (Sibghat-Ullah & Day, 1992), 6-thioguanine:T pairs (Griffin et al, 1994), DiAP:T pairs, AMAP:T mispairs (Sibghat-Ullah & Day, 1995b), and G:U mispairs (Neddermann & Jiricny, 1994). With each of these substrates, the product displayed the characteristic features of short-patch repair: incision was observed on only the strand containing the mismatched T, leading to the production of a one nucleotide gap at that site. Of interest was the fact that the G:T repair activity in the cell-free extract recognized mismatches with a variety of base pair structures in DNA: the G:T pair is reported to be in wobble configuration (Kennard, 1985; Swann, 1990), the DiAP:T pair is a strong Watson−Crick pair (Coll et al., 1986; Chazin et al., 1991), and the m6G:T (Kalnick et al., 1989; Swann, 1990) and AMAP:T pairs (by analogy to m6G:T) are weak Watson−Crick pairs. It was thus of interest to determine whether the recently-purified thymine glycosylase (Neddermann & Jiricny, 1993) would follow the substrate recognition pattern of the activity found in the cell-free extract.

Using recombinant human G:T specific thymine glycosylase, we assayed the *in vitro* incision of DNA substrates containing G:T and modified base mismatches at defined sites in a 45-bp ds DNA (Figure 1A). We also assayed the incision of DNAs containing G:T mispairs at the same site in four contexts, ApG, CpG, TpG, and GpG. Our results show that the recombinant human enzyme incises the DNA substrate having a single G:T mispair within a CpG sequence with high specificity. It incises G:T mispairs in ApG, TpG, and GpG contexts much less efficiently and shows weak activity toward m6G:T and AMAP:T mispairs. In contrast to the A1235 cell extracts, the purified enzyme failed to process the DiAP:T heteroduplex.

MATERIALS AND METHODS

Enzymes

G:T mismatch-specific thymine−DNA glycosylase was expressed in *E. coli* and purified to apparent homogeneity by a series of chromatographic steps (Neddermann et al., 1996). The pure, recombinant protein (1 μg/μL) was used in this study. T4 kinase was from Pharmacia, and T4 DNA ligase was from BRL.

DNA Substrates

All oligonucleotides that did not contain base analogs were synthesized by the DNA Synthesis Laboratory, University of Alberta. Oligos containing m6G were synthesized by the Regional DNA Synthesis Laboratory, University of Calgary. Oligos containing DiAP, AMAP, or m4T were prepared in the laboratory of Y.-Z. Xu as described (Xu et al., 1992a,b). The oligo containing 2AP was prepared in the laboratory of M. F. Goodman (Bloom et al., 1993). The substrates were labeled at the 5′ end with T4 kinase and [γ-^{32}P]ATP using the appropriate single-stranded oligo before preparing the 45-bp DNA duplexes as described (Sibghat-Ullah & Day, 1992).

Incision Assay

The incision assay (Wiebauer & Jiricny, 1989; Sibghat-Ullah & Day, 1992) was performed as follows: Reaction mixtures (50 μL) contained 1 μL of diluted recombinant G:T mismatch glycosylase (stock protein concentration 1 μg/μL; diluted 1/5, 1/10, 1/20, or 1/40 prior to pipetting the 1 μL) and 1 ng (33 fmol) of DNA substrate (5′-terminally labeled with ^{32}P in the strand containing the mismatched T) in 25 mM Hepes, pH 7.9, 50 mM KCl, 1 mM DTT, 0.5 mM EDTA, 0.01 mM $ZnCl_2$ and 0.1 mg/mL BSA. Reactions were incubated at 30 °C for 0−100 min followed by extraction with phenol and chloroform. NaOH was added to the aqueous layers to a final concentration of 0.1 N, and the DNAs were heated at 90 °C for 30 min. DNA was then precipitated by ethanol, dried in a Speedvac, and dissolved in formamide/dye. Incision products were analyzed by electrophoresis on 12% sequencing gels followed by autoradiography on X-ray film (Sibghat-Ullah & Day, 1992).

The extent of the incision reaction was quantitated by scanning the autoradiograms using an LKB laser densitometer and XGELSCAN software as before (Sibghat-Ullah & Day, 1995b).

RESULTS

Incision at G:T and Modified G:T Mispairs in the DNA

The structure of the basic DNA substrate is shown in Figure 1A. The G:T mispair at site 25 was replaced with m6G:T, DiAP:T, AMAP:T, G:m4T, and 2AP:T mispairs (the purines always replacing the G) in the other substrates (Figure 1B). Increasing amounts (25, 50, 100, and 200 ng) of recombinant enzyme were assayed for incision with 33 fmol (1 ng) of a 45-bp DNA containing G:T mispairs and 5′-terminally labeled in the strand containing the mismatched T, under standard reaction conditions (Sibghat-Ullah & Day, 1992), and the 100 ng amount was selected for kinetic studies. The rate of the incision of the G:T-containing 45-bp duplex was then compared with the rates of cleavage of the other substrates, similarly labelled at the 5′ termini of their T strands. Figure 2 (lanes 1−4) and Figure 3 (lanes 1−3) show that under standard reaction conditions, the 45-bp DNA containing a G:T mispair was converted quantitatively to a product 20 nt in length in 40 min. Under similar conditions, the 45-bp DNAs containing m6G:T and AMAP:T mispairs were processed by the enzyme to give rise to

A

```
5' GGCCAGCTAGTGGTGGTGGGCGCCGGCGGTGTGGGCATTCGTAAT 3'
3' CCGGTCGATCACCACCACCCGCGGTCGCCACACCCGTAAGCATTA 5'
```

B

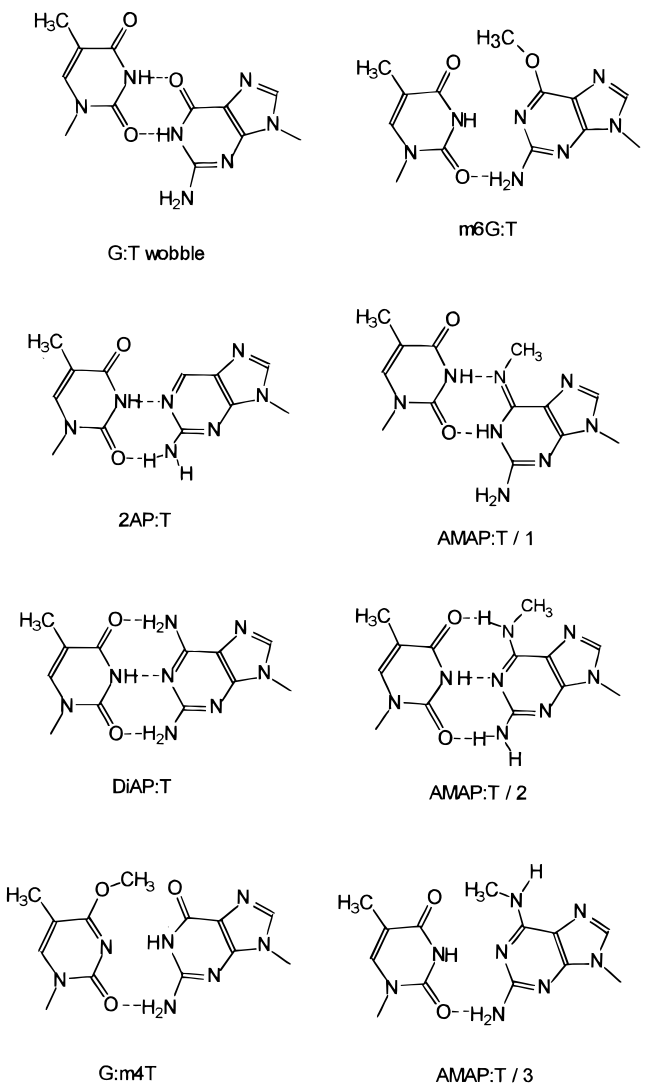

FIGURE 1: (A) Structure of the 45-bp DNA substrates. The G:T base mismatch is located at site 25 in the DNA. DNAs containing DiAP:T, m6G:T, G:m4T, 2AP:T, and AMAP:T pairs at site 25 in the DNA were also studied. In each case the purine of the pair was in the top strand (the strand containing the mismatched G) and the thymine (or modified thymine) in the bottom strand. All substrates were prepared as described previously (Sibghat-Ullah & Day, 1992, 1993, 1995a,b). (B) Proposed structures of the mismatched base pairs employed in this study. Evidence in support of the structures is presented in the following references: G:T (Brown et al., 1985; Kennard, 1985); m6G:T (Kalnick et al., 1989; Swann, 1990); G:m4T (Kalnick et al., 1988); DiAP:T (Coll et al., 1986; Chazin et al., 1991); 2AP:T (Sowers et al., 1986). To our knowledge the structure of the AMAP:T pair has not been studied. It may most closely be approximated by that of the m6G:T pair as in AMAP:T/3.

significant amounts of 20 nt incision fragments (Figure 2, lanes 8−10; Figure 3, lanes 7−9). By contrast, incubation with the enzyme of the 45-bp DNAs having DiAP:T mispairs (Figure 3, lanes 4−6), 2AP:T mispairs (Figure 2, lanes 11−13), or G:m4T mispairs (Figure 2, lanes 5−7) produced some product, but no more than that produced by incubation with fully base-paired DNA (Figure 2, lanes 14−16; Figure 3, lanes 10−12). In other experiments, incubation of DNAs with DiAP:T, 2AP:T, G:m4T, or G:C pairs with enzyme produced no detectable product (e.g., Figure 7, lane 12 and data not shown).

We quantified the 45-nt substrate bands and the 20-nt product bands in Figures 2 and 3 to determine the rates of enzymatic incision of the various substrates. The results are shown in Figure 4. The amount of 20-nt fragment generated from the DNA containing the G:T mispair increased linearly between 0 and 40 min with a rate of 0.7 fmol/min, at which time all of the substrate was consumed. The DNAs having m6G:T and AMAP:T pairs yielded 20-nt product at lower rates: 0.38 and 0.15 fmol/min, respectively. The percentages of substrate converted to 20-nt product were 40% and 20% at 40 min and did not increase thereafter (Figure 2; other data not shown), possibly due to inactivation of the enzyme. Treatment of DNAs containing DiAP:T, 2AP:T, and G:m4T pairs generated ~1% product at 40 min, much the same as the control DNA with a G:C pair. We conclude that while the recombinant enzyme recognizes the G:T mismatched pair well, it has limited ability to recognize pairs between G analogs and T in DNA by contrast with the action of a cell-free extract that had activity on G:T, m6G:T, DiAP:T, and AMAP:T pairs, but not on G:m4T pairs. (The 2AP:T pair was not tested in the former study.)

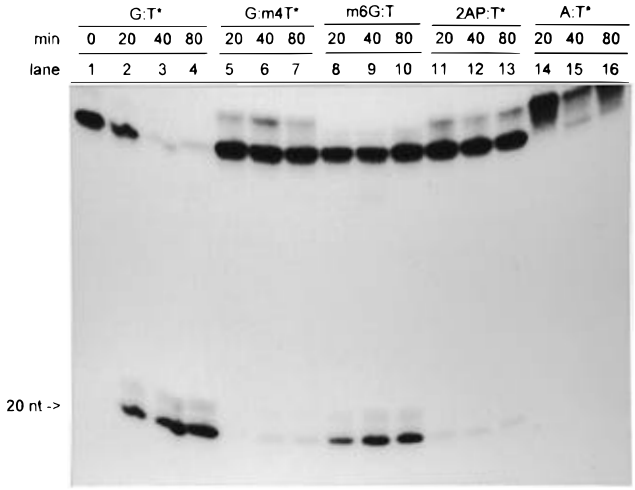

FIGURE 2: Incision of 45-bp DNAs with G:T and base analog pairs. DNA substrates (33 fmol), ^{32}P labeled in the 5′ terminus of the strands containing mismatched T, were incubated with 100 ng of thymine glycosylase in 50 μL reaction volumes containing 25 mM Hepes, pH 7.9, 50 mM KCl, 1 mM DTT, 0.01 mM $ZnCl_2$, 0.5 mM EDTA, and 100 μg of BSA/mL. Reactions were allowed to proceed for the specified times. The DNAs were processed to break phosphodiester bonds at the AP sites created by the glycosylase and subjected to electrophoresis as described under Materials and Methods. DNA substrates, represented by the base pairs they contain at position 25, are designated G:T in lanes 1−4, G:m4T in lanes 5−7, m6G:T in lanes 8−10, 2AP:T in lanes 11−13, and A:T in lanes 14−16.

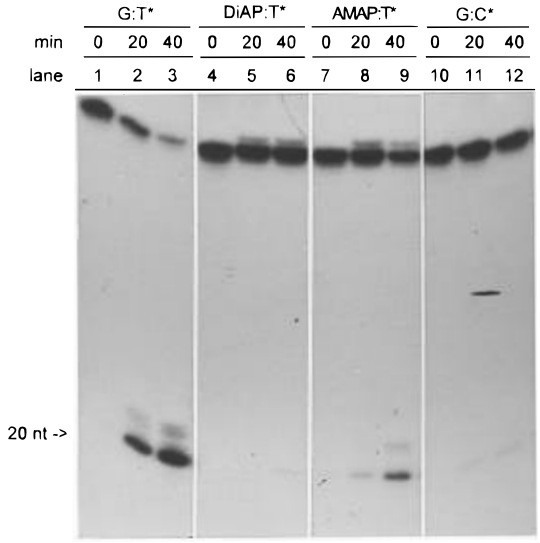

FIGURE 3: Incision of 45-bp DNA substrates with G:T and base analog pairs. Reactions were performed as described in the legend to Figure 2. The various substrates are designated as G:T DNA (lanes 1−3), DiAP:T DNA (lanes 4−6), AMAP:T DNA (lanes 7−9), and G:C DNA (lanes 10−12).

Effect on the Mismatch Glycosylase of the Base 5′ to the G:T Mispair

The C:G base pair immediately to the left of the G:T mispair, boxed along with the G:T mismatch in the DNA in Figure 1A, was replaced with A:T, T:A, or G:C pairs to create 45-bp DNA substrates having single G:T mispairs in four different contexts, 5′-ApG-3′/3′-TpT-5′, 5′-GpG-3′/3′-CpT-5′, 5′-TpG-3′/3′-ApT-5′, and 5′-CpG-3′/3′-GpT-5′ (Sibghat-Ullah & Day, 1995a). All four DNA substrates were 5′-terminally labeled in the strand containing the mismatched T, and ∼1 ng of each was then incubated with the glycosylase and assayed for product formation. The appearance of 20-nt fragments in Figure 5, lanes 1−4, indicated that the glycosylase removed the mismatched thymine from all mismatched DNAs. The 20-nt fragments were nearly absent from the eight control DNAs that lacked mismatches (lanes 5−12), showing that the product formation in each lane (1−4) depended on the presence of the base mismatch. The extent of product formation depended on the base 5′ to the G:T mispair in the DNA. We quantified the amounts of 20-nt product in lanes 1−4 and present them in Figure 6 as

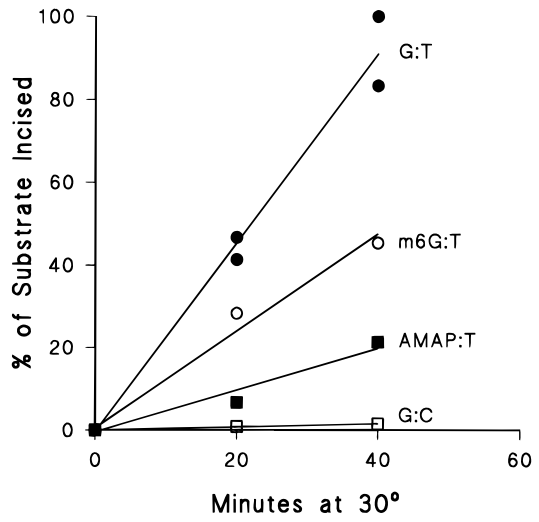

FIGURE 4: Rates of reaction between recombinant mismatch glycosylase and correctly-paired or aberrantly-paired DNAs. Quantitation of the incision product was achieved by scanning the 20-nt bands shown in Figures 2 and 3 as described under Materials and Methods. The amount of substrate (33 fmol at zero time) converted to product is shown at different time points. Base pairs indicate curves for the different substrates. The amount of product did not increase after 40 min of incubation, probably due to the inactivation of the enzyme.

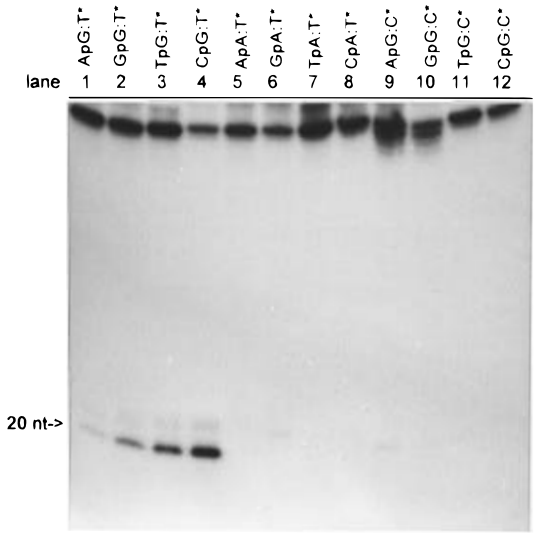

FIGURE 5: Effect of the neighboring base 5′ to the G:T mismatch in the substrate on the glycosylase activity. The 45-bp DNA substrates containing G:T in four different sequence contexts (5′-CpG-3′/3′-GpT-5′, 5′-ApG-3′/3′TpT-5′, 5′-GpG-3′/3′-CpT-5′, and 5′-TpG-3′/3′-ApT-5′) and control DNAs were prepared with label in the 5′ termini of their bottom strands. 33 fmol of each substrate was incubated in 90 min reactions with the enzyme as described in Figure 2. Lanes are designated NpG:T, NpA:T, or NpG:C according to the DNA in the reaction. (Note that the enzyme yielded somewhat less product from the CpG:T DNA than in Figures 2−3).

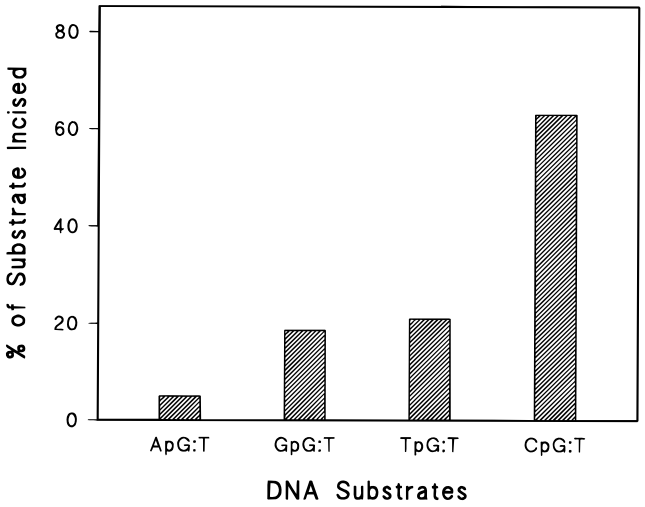

FIGURE 6: Comparison of amounts of product obtained with DNAs of different NpG:T sequences. The incision products obtained in Figure 5 were quantitated by scanning the bands as described in Materials and Methods.

percent of the substrate incised. 64% of the DNA with the CpG:T mismatch was converted to product as compared with 21% of the TpG:T DNA, 19% of the GpG:T DNA, and 5% of the ApG:T DNA. Changing the base 5′ to a G:T mismatch at site 20 (only data for G:T mismatches at site 25 are shown in Figure 5) caused a similar variation in the amount of product formation (data not shown).

Sequence Effects on m6G:T Mismatch Repair in DNA

To study the effect of neighboring bases on the enzyme's action on DNAs containing m6G:T pairs, we prepared DNAs with m6G substituted for G at site 25 in the four sequence contexts and processed them as above. Figure 7 shows that the enzyme produced detectable product only when the m6G:T pair was preceded by a pyrimidine (lanes 3−4). In this respect the recombinant enzyme behaved much like the cell-free extracts used earlier (Sibghat-Ullah & Day, 1995a). The enzyme failed to produce product from DNAs that have A or G 5′ to the m6G:T pair (lanes 1−2). This finding lends support to the idea that the enzyme generally prefers m6G:T or G:T mispairs present in the CpG:T or Cpm6G:T context (and to varying extents when present as Tpm6G:T or TpG:T).

Lack of Incision at Guanine:5-Methylcytosine Pairs

Jost et al. (1995) characterized a 52.5-kDa 5-methylcytosine DNA glycosylase from 12-day chick embryos. The enzyme may also have mismatch-specific thymine glycosylase activity because the G:m5C and G:T glycosylase activities copurified. Thus our human recombinant enzyme may also have G:m5C glycosylase activity. To investigate this possibility, we compared the activity of the recombinant G:T mismatch-specific thymine DNA glycosylase on DNAs prepared with CpG:m5C, CpG:T, or CpG:U pairs at site 25. Figure 7 (lanes 9−13) shows that the enzyme has an undetectable level of activity on CpG:m5C pairs (lane 9) but has activity on CpG:T (lane 11) and CpG:U (lane 13) mismatched pairs. The activity on CpG:U pairs has been observed in another substrate previously (Neddermann & Jiricny, 1994).

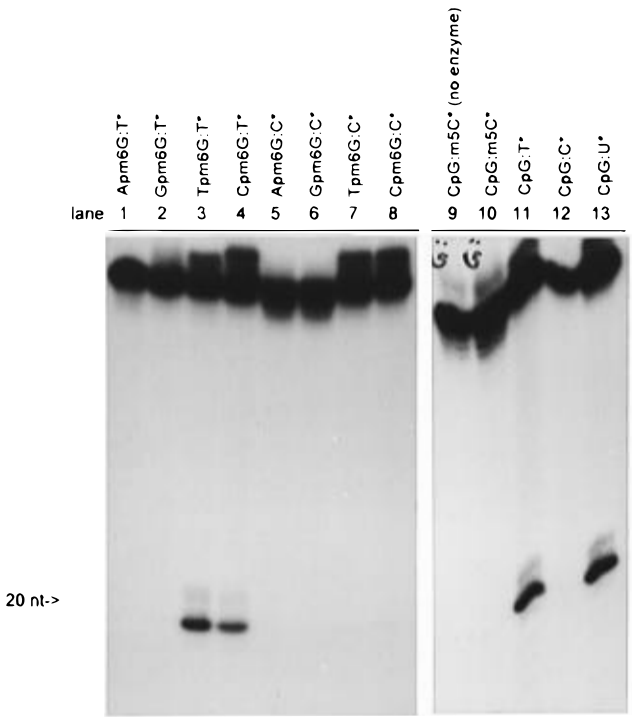

FIGURE 7: Effect of the neighboring base 5′ to the m6G:T pair in the substrate on the glycosylase activity and lack of incision at G:m5C pairs. Lanes 1−8: DNA substrates with the general form Npm6G:T (33 fmol) were incubated with the enzyme for 90 min. DNAs were treated to break phosphodiester bonds at AP sites and subjected to electrophoresis. (Note that the 90 min incubation of enzyme with the Cpm6G:T substrate yielded somewhat less product than the reaction analyzed in Figure 2). Lanes 9−13: DNA substrates of the general form CpG:Py (Py = m5C, T, U, and C) were incubated with the enzyme for 90 min and analyzed as above.

DISCUSSION

The recombinant human mismatch glycosylase acted on DNA mispairs in the order G:T > m6G:T > AMAP:T > DiAP:T ∼ G:m4T ∼2AP:T ∼ G:C ∼ A:T ∼ G:m5C. It preferred CpG:T mispairs to TpG:T, ApG:T, or GpG:T mispairs at two DNA sites. It preferred Cpm6G:T over Tpm6G:T pairs. Activity on Gpm6G:T and Apm6G:T pairs was not detected.

In our earlier study (Sibghat-Ullah & Day, 1995a), we used the present series of oligonucleotides (with the exception of the 2AP:T DNA) to investigate the sequence dependence and substrate specificity of human mismatch-specific thymine DNA glycosylase activity present in unfractionated A1235 extracts. Our data showed that, with the exception of the matched substrates G:C and A:T as well as the G:m4T mispair, the substrates G:T, DiAP:T, m6G:T, and AMAP:T were incised at the site of the mispaired T or purine-analog paired T, albeit with varying efficiencies. We used the purified, recombinant human G/T mismatch glycosylase to determine whether all the structures might be addressed by the same enzyme. Significantly, the pure enzyme failed to address the DiAP:T mispair, implying that it was a substrate for a distinct as yet unidentified glycosylase or endonuclease in the extract used (Sibghat-Ullah & Day, 1995). Possibilities for DiAP:T activity include newly reported glycosylases, the substrates of which range from "simple" mismatches (McGoldrick et al., 1995) to DNA bases modified by vinyl chloride (Dosanjh et al.,1994) or by benzoquinone (Chenna et al., 1995), as well as enzymes (e.g., topoisomerase I) that

have been reported to possess mismatch-specific endonuclease activities (Yeh et al., 1994).

The fact that only the G:T, m6G:T, and AMAP:T substrates were incised well by the purified G/T glycosylase leads us to speculate on the structural features required by the enzyme for substrate recognition. As shown in Figure 1B [see also Kennard (1985), Swann (1990), Coll et al. (1986), Chazin et al. (1991), and Kalnick et al. (1990)] the substrates fall into three structural classes. Members of the first class, the G:T (and G:U) mispairs, are stabilized within the DNA helix by two hydrogen bonds. With reference to Watson−Crick A:T or G:C pairs, the constituent bases of G:T or G:U pairs are tilted in opposite directions, such that the purine is displaced somewhat into the minor groove and the pyrimidine is displaced somewhat into the major groove to form a wobble-base pair. Apart from forming a rigid distortion in the helix, this alteration almost certainly affects the stacking energy.

The second structural class contains the G:C and A:T pairs, as well as the 2AP:T and DiAP:T mispairs, none of which has the helix-distorting features of the former class. The analog pair structures would be expected to resemble Watson−Crick pairs with two or three hydrogen bonds (Figure 1B), and, correspondingly, they ought not be substrates for the mismatch-specific G:T glycosylase. Our data agree with this idea.

The third structural class includes the m6G:T, m4T:G, and AMAP:T mispairs. In a way similar to the Watson−Crick base pairs, these structures would be expected neither to generate much distortion of the DNA helix nor to be incised by the G:T glycosylase. However, the two of these three structures that possess an unmodified thymine are substrates for the enzyme. We offer two explanations. In the first, the hydrophobic alkyl substituent may alter the solvation of the DNA molecule such that the DNA might be distorted at the site of the mismatch in a way similar to the distortion at the G:T mispair. The second explanation is that the alkylation of the bases reduces the stability of the "base pairs": let us assume that the G:T glycosylase functions by a mechanism similar to the uracil DNA glycosylase (Mol et al., 1995; Savva et al., 1995) and that it "flips" the thymine out of the helix prior to cleaving the base from the sugar-phosphate backbone. In this case, the weaker the base-pairing, the less the energy required for flipping-out. Thus, the AMAP:T mispair, which might form three hydrogen bonds, could be more stable and less likely to be processed by the glycosylase than the m6G:T mispair, which forms one hydrogen bond (Swann, 1990; Kalnick et al. 1989). As shown in Figures 2 and 3, this expectation is borne out. On the basis of these data, it might have been expected that the m4T:G pair, which also forms one hydrogen bond (Kalnick et al., 1988), would be processed with an efficiency similar to that of the m6G:T mismatch. However, the latter mispair differs from the others in that it contains a modified thymine residue, which may not fit into the active site of the glycosylase. The m4T:G substrate may be poorly processed for this reason alone.

The second part of the study regarded the effects of immediate flanking sequences on the efficiency of G:T mismatch processing. The results obtained with the purified glycosylase showed that, at two DNA sites, the preferred substrate was a G:T mispair in the context of a CpG dinucleotide. This was distinct from an activity in extracts that incised G:T mismatches irrespective of the base 5′ to the mismatched G (Sibghat-Ullah & Day, 1995a). Presuming that the G:T glycosylase evolved to correct G:T mispairs arising by the spontaneous hydrolytic deamination of 5-methylcytosine and given that cytosine methylation in vertebrates is found predominantly in the CpG sequence context, the observed substrate preference would be expected. However, as the difference between the best and worst sequence contexts is only 12-fold, it cannot be assumed that the enzyme would not address a G:T mispair in, for example, an ApG context. Indeed, our SV40 transfection experiments demonstrated that G:T mismatches in CpG and ApG contexts were repaired with equal efficiency during the time between transfection and the first round of SV40 replication (Brown & Jiricny, 1987). Our current data show that the enzyme may attempt to correct a subset of m6G:T mispairs generated by chemical modification of DNA. O^6-Methylguanine can be produced in any sequence context and, if unrepaired, will be base-paired either with C prior to replication or with T after replication [see Karran and Bignami (1992) for review]. Both m6G:C and m6G:T are addressed by the post-replicative mismatch correction pathway. Indeed, the long-patch mismatch repair pathway is responsible for cytotoxicity caused by m6G in cells deficient in MGMT (Aquilina et al., 1995). Should the m6G-containing base pairs escape correction by the long-patch repair mechanism, some would remain as substrates for the G:T glycosylase. However, because alkylation-sensitive cells (deficient in MGMT) can escape m6G-induced killing if they harbor a mutation that inactivates the long-patch repair pathway [see also Branch et al. (1995)], any contribution of the glycosylase to stimulating m6G-induced killing is likely to be small.

The G:T glycosylase may interfere with long-patch mismatch repair. Given that the G:T glycosylase addresses G:T mispairs outside the CpG context (Figure 5), the enzyme might attempt to correct G:T mispairs that arise as errors of DNA replication. Such interference would lead to mutations, as all G:T mismatches would be repaired to G:C, irrespective of any strand discrimination signals that dictate repair of G:T to A:T. However, our data (P. Gallinari and J. Jiricny, unpublished) indicate that this would be rather unlikely, as the mismatch binding factor of the long-patch mismatch repair system is very abundant and binds to G:T mispairs with an affinity significantly greater than the G:T glycosylase.

In summary, our data suggest that the best substrate of the mismatch-specific thymine DNA glycosylase is the G:T mispair in a CpG sequence context. This would imply that the enzyme has indeed evolved to correct G:T mispairs arising as the result of spontaneous hydrolytic deamination of 5-methylcytosine in CpG dinucleotides.

ACKNOWLEDGMENT

We thank Halyna Marusyk for expert preparation of the figures.

REFERENCES

Aquilina, G., Hess, P., Fiumicino, S., Ceccotti, S., & Bignami, M. (1995) *Cancer Res. 55*, 2569−2575.

Bloom, L. B., Otto, M. R., Beechem, J. M., & Goodman, M. F. (1993) *Biochemistry 32*, 11247−11258.

Branch, P., Hampson, R., & Karran, P. (1995) *Cancer Res. 55*, 2304−2309.

Brown, T., Kennard, O., Kneale, G., & Rabinowich, D. (1985) *Nature 315*, 604−606.
Brown, T. C., & Jiricny, J. (1987) *Cell 50*, 945−950.
Chazin, W. J., Rance, M., Chollet, A., & Leupin, W. (1991) *Nucleic Acids Res. 19*, 5507−5513.
Chenna, A., Hang, B., Rydberg, B., Kim, E. Pongracz, K., Bodell, W. J., & Singer, B. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 5890−5894.
Claverys, J.-P., & Lacks, S. A. (1986) *Microbiol. Rev. 50*, 133−165
Coll, M., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H., & Rich, A. (1986) *J. Biomol. Struct. Dyn. 4*, 157−172.
Dosanjh, M. K., Chenna, A., Kim, E., Fraenkel-Conrat, H., Samson, L., & Singer, B. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 1024−1028.
Duncan, B. K., & Miller, J. H. (1980) *Nature 287*, 560−561.
Frederico, L. A., Kunkel, T. A., & Shaw, B. R. (1990) *Biochemistry 29*, 2532−2537.
Griffin, S., & Karran, P. (1993) *Biochemistry 32*, 13032−13039.
Griffin, S., Branch, P., Xu, Y.-Z., & Karran, P. (1994) *Biochemistry 33*, 4787−4793.
Hennecke, F., Kolmar, H., Bründl, K., & Fritz, H.-J. (1991) *Nature 353*, 776−778.
Holmes, J., Jr., Clark, S., & Modrich, P. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 5837−5841.
Jost, J.-P., Siegmann, M., Sun, L., & Leung, R. (1995) *J. Biol. Chem. 270*, 9734−9739.
Kalnick, M. W., Kouchakdjian, M., Li, B. F. L., Swann, P. F., & Patel, D. J. (1988) *Biochemistry 27*, 108−115.
Kalnick, J. W., Li, B. F. L., Swann, P. F., & Patel, D. J. (1989) *Biochemistry 28*, 6170−6181, 6182−6192.
Karran, P., & Bignami, M. (1992) *Nucleic Acids Res. 20*, 2933−2940.
Kennard, O. (1985) *J. Biomol. Struct. Dyn. 3*, 205−225.
Levin, J. D., & Demple, B. (1990) *Nucleic Acids Res. 18*, 5069−5075.
McGoldrick, J. P., Yeh, Y.-C., Solomon, M., Essigmann, J. M., & Lu, A.-L. (1995) *Mol. Cell. Biol. 15*, 989−996.
Modrich, P. (1991) *Annu. Rev. Genet. 25*, 229−253.
Modrich, P. (1994) *Science 266*, 1959−1960.
Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E., & Tainer, J. A. (1995) *Cell 80*, 869−878.
Neddermann, P., & Jiricny, J. (1993) *J. Biol. Chem. 268*, 21218−21224.
Neddermann, P., & Jiricny, J. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 1642−1646.
Neddermann, P., Gallinari, P., Lettieri, T., Schmid, D., Truong, O., Hsuan, J. J., Wiebauer, K., & Jiricny, J. (1996) *J. Biol. Chem. 271*, 12767−12774.
Radman, M. (1988) in *Genetic Recombination*, Kucherlapati, R., & Smith, G. R. Eds.) pp 169−192, American Society for Microbiology, Washington, DC.
Razin, A., & Riggs, A. D. (1980) *Science 210*, 604−610.
Savva, R., McAuley-Hecht, K., Brown, T., & Pearl, L. (1995) *Nature 373*, 487−493.
Shapiro, R., & Klein, R. S. (1966) *Biochemistry 5*, 2358−2361.
Sibghat-Ullah, & Day, R. S., III (1992) *Biochemistry 31*, 7998−8008.
Sibghat-Ullah, & Day, R. S., III (1993) *Nucleic Acids Res. 21*, 1281−1287.
Sibghat-Ullah, & Day, R. S., III (1995a) *Biochemistry 34*, 6869−6875.
Sibghat-Ullah, & Day, R. S., III (1995b) *Biochemistry 34*, 7438−7442.
Sowers, L. C., Fazakerley, G. V., Eritja, R., Kaplan, B. E., & Goodman, M. F. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5434−5438.
Swann, P. F. (1990) *Mutat. Res. 233*, 81−94.
Thomas, D. C., Roberts, J. D., & Kunkel, T. A. (1991) *J. Biol. Chem. 266*, 3744−3751.
Wiebauer, K., & Jiricny, J. (1989) *Nature 339*, 234−236.
Wiebauer, K., & Jiricny, J. (1990) *Proc. Nat. Acad. Sci. U.S.A. 87*, 5842−5845.
Xu, Y.-Z., Zheng, Q., & Swann, P. F. (1992a) *Tetrahedron 48*, 1729−1740.
Xu, Y.-Z., Zheng, Q., & Swann, P. F. (1992b) *J. Org. Chem. 57*, 3839−3845.
Yeh, Y.-C., Lu, H.-F., Ellis, C. A., & Lu, A.-L. (1994) *J. Biol. Chem. 269*, 15498−15504.

BI961022U