DNA-substrate sequence specificity of human G:T mismatch repair activity

Sibghat-Ullah and Rufus S.Day, III*

Molecular Oncology Program, Department of Medicine, Cross Cancer Institute, 11560 University Avenue, Edmonton, Alberta T6G 1Z2, Canada

Received October 6, 1992; Revised and Accepted January 26, 1993

ABSTRACT

G:T mispairs in DNA originate spontaneously via deamination of 5-methylcytosine. Such mispairs are restored to normal G:C pairs by both *E.coli* K strains and human cells. In this study we have analyzed the repair by human cell extracts of G:T mismatches in various DNA contexts. We performed two sets of experiments. In the first, repair was sequence specific in that G:T mispairs at CpG sites at four different CpG sites were repaired, but a G:T mismatch at a GpG site was not. Cytosine hemimethylation did not block repair of a substrate containing a CpG/GpT mismatch. In the second set of experiments, substrates with a G:T mismatch at a fixed position were constructed with an A, T, G, or C 5' to the mismatched G, and alterations in the complementary strand to allow otherwise perfect Watson – Crick pairing. All were incised just 5' to the mismatched T and competed for repair incision with a G:T substrate in which a C was 5' to the mismatched G. Thus human G:T mismatch activity shows sequence specificity, incising G:T mismatched pairs at some DNA sites, but not at others. At an incisable site, however, incision is little influenced by the base 5' to the mismatched G.

INTRODUCTION

G:T base mismatches are produced in DNA by the spontaneous hydrolytic deamination of 5-methylcytosine (1,2). Enzymatic activities in human cell-free extracts are capable of incising the phosphodiester bonds immediately 5' and 3' to the mismatched T (3–6), generating a one nucleotide gap; the strand containing the mismatched G is not severed. In the presence of Mg^{++} , ATP and the 4 DNA-dNTPs the G:T mispair is corrected exclusively to G:C pairs (3,4), in accordance with the results of *in vivo* mismatch repair studies (7). Wiebauer and Jiricny (4) concluded that one of the enzymes constituting the activity acts as a thymine-specific glycosylase because free thymine is generated during the processing of the mismatch. A DNA substrate containing a matched A:T base pair in place of the mismatched G:T pair failed to serve as a substrate for the generation of free thymine in this reaction. The AP

endonuclease(s) or terminus-processing enzymes (5'-deoxyribosephospho- diesterases [8] or 3'-'repair' diesterases [9]) involved in nicking and preparing the phosphodiester backbone for further repair have not been characterized. However, the one nucleotide gap is filled in by polymerase β (because anti-polymerase β antibodies block the reaction) and the strand is ligated, completing the process of restoration of the G:C base pair to the DNA (4).

Several considerations led us to seek evidence for involvement of DNA sequence-specificity in human G:T mismatch repair. DNA sequence specificity of G:T mismatch repair was noted (6) as one possible explanation for ability of some workers to observe G:T mismatch incision of substrates containing O⁶-methylguanine (m6G):T mismatches (6) and inability of others (10,11). Further, in *E.coli* K strains, the Dcm (deoxycytidine methylation) protein methylates the 3' cytosine of the sequence 5'-CC(A/T)GG-3' (12) and the Vsr (very short patch repair) G:T mismatch repair system restores 5'-CC(A/T)G-3' or 5'-C(A/T)GG-3' sequences in a mechanism involving PolII, a short patch DNA polymerase (12–15). This consideration led to the possibility that mammalian cells may perform G:T-specific mismatch repair, also involving a short patch polymerase, in a DNA sequence-specific fashion, perhaps in order to restore the CpG sequences at which cytosine is methylated by mammalian systems (16). In this regard, it was interesting to note that the mismatched G's in the substrates used previously (3–6) in which G:T-specific mismatch repair of the kind reported by Wiebauer and Jiricny (3,4) was observed, were contained in CpG sequences. Studies in which human cell extracts were not reported to contain the type of G:T-specific mismatch repair activity reported by Wiebauer and Jiricny (3,4), but in which G:T base mismatches were repaired, employed substrates in which the mismatched G was not preceded by a C. The conclusions of these latter studies were (i) that the nicked G-containing strand of a plasmid containing a G:T mismatch was preferentially repaired (17) and (ii) that G:T mismatches were repaired by a long patch repair system that was blocked by inhibitors of DNA polymerase α (18). However, a study done partly *in vivo*, not with extracts, and in which factors such as recombination or non-specific exonuclease action followed by DNA synthesis may have masqueraded as mismatch repair,

* To whom correspondence should be addressed

showed restoration of a G:T mismatch in which the mismatched G was 3' to an A (7).

In this paper we present a study of the effect of the environment at the site of the G:T mismatch on the repair of G:T mismatches using a cell-free system (6,19,20). As shown by incision, substrate-competition and repair synthesis assays, DNA substrates having single G:T base mismatches at CpG dinucleotides served as substrates for G:T mismatch repair, whereas a substrate with a G:T mismatch at a GpG dinucleotide did not. However, in a second set of experiments, changing the CpG (in which the G was the mismatched G at a fixed site in the substrate) to an ApG, TpG or GpG in one strand (with alterations in the complementary strand to accommodate Watson-Crick pairing at all sites but that of the G:T mismatch) did not greatly influence mismatch incision.

MATERIALS AND METHODS

Cell-free extracts and enzymes

Cell line A1235 (6; a generous gift from Dr S.A. Aaronson, NCI, Bethesda, MD) was propagated in DMEM supplemented with 5% fetal calf serum. Cell-free extracts were prepared by the method of Manley et al. (19) as previously described (6,20) and frozen in aliquots at -85°C . HpaII, NarI, HhaI and T4 polynucleotide kinase were purchased from BRL, USA. [γ - ^{32}P] ATP (5,000 Ci/mmol) was from Amersham.

Oligodeoxynucleotides, preparation of DNA duplexes and labelling

All oligodeoxynucleotides (oligos; 45-nt) used in this study (see Figs. 1 and 5) were prepared by the DNA Synthesis Laboratory, Department of Microbiology, University of Alberta by conventional β -cyanoethyl phosphoramidite techniques. Individual oligos were purified on 12% denaturing acrylamide gels; DNA was extracted from gel slices by electroelution. The oligos were labelled in their 5' termini by T4 kinase and [γ - ^{32}P] ATP. Substrates were prepared by hybridizing appropriate strands as described previously (6). Concentrations of unlabelled oligonucleotides were determined by 260 nm spectrophotometry; concentrations of labelled substrates are based on an assumed 100% recovery from the gel-purification and ethanol precipitation steps and are approximations.

Incision assay

The mismatch incision assay was performed as described previously (6). DNA duplexes were incubated with cell-free extract in 25 mM Hepes, pH 7.9, 0.5 mM DTT, 0.5 mM EDTA, 0.01 mM ZnCl_2 , and 100 $\mu\text{g/ml}$ BSA in 50 μl volumes for 12 hours at 30°C . To perform competition experiments, DNA substrates were radioactively labelled; competing oligos were unlabelled. Following incubation of extract with 2 ng of substrates and 0–20 ng of competing DNAs, reaction mixtures were extracted with phenol-chloroform and precipitated with ethanol. Reaction products were analyzed by electrophoresis on 12% sequencing gels followed by autoradiography (6). Details concerning individual experiments are presented in the text or the figure legends.

Repair synthesis assay

Repair synthesis following the incision of DNAs containing G:T base mismatches was assayed by the ability of the cell-free extract to correct the G:T-modified restriction site to a G:C-containing

site by a method similar to that of Wiebauer & Jiricny (3). DNA substrates were incubated with cell-free extract in 25 mM Hepes buffer, pH 7.9, 40 mM KCl, 0.5 mM EDTA, 1 mM MgCl_2 , 50 μM ATP, 30 μg phospho(enol)pyruvate, 1 U pyruvate kinase, and 200 μM of each dNTP in 50 μl volumes. Reactions were for 12 hours at 30°C . Following extraction as above, the recovered DNA was digested with the appropriate restriction enzyme for 1 h at 37°C . Corresponding restriction fragments were identified by electrophoresis of the DNA digest through 12% sequencing gels followed by autoradiography at -80°C for 24 hours.

RESULTS

Incision of 45-bp DNAs containing single G:T mismatches at defined locations

To determine the effect on G:T mismatch incision activity of altering the DNA sequence surrounding the G:T base mismatch, we constructed the 45-bp DNA substrates shown in Fig. 1. The substrate DNAs in Fig. 1A–1E contain single G:T mispairs (enclosed by boxes) within restriction endonuclease recognition sequences. In all cases the mismatch blocks digestion by the designated restriction enzyme. Four of the 45-bp DNAs (Fig. 1A, 1C, 1D and 1E) contain single G:T mispairs at a CpG dinucleotide sequence. The sequence of the DNA in Fig. 1A, from the M13mp9-based 90 nt substrate in refs. 3 and 4, was much different from remainder of the DNAs. The DNA sequences Fig. 1B–G are one base-variations of a sequence from the H-ras gene sequence (refs. 6 and 21): the sequence of Fig. 1B contains a G:T mispair in which a G is 5' to the mismatched G; the sequences in Figs. 1C and 1D have the same top strands as that in Fig. 1B but the G:T mismatches are at different sites and are preceded 5' by one C (Fig. 1C) or 2 C's (Fig. 1D). The substrate shown in Fig. 1E is the same as that in Fig. 1D, but the C that is 5' to the mismatched G is 5-methylated. The substrates in Fig. 1F and Fig. 1G contain no mismatches but the former is fully methylated at the C that is hemimethylated in Fig. 1E, and the latter is non-methylated.

Fig. 2 presents the results of the *in vitro* G:T mismatch incision assay obtained after incubating cell-free extract from A1235 cells with mismatched 45-bp substrates (Figs. 1A, 1B, 1C, 1D, 1E) and control substrates (Fig. 1G and one similar to that in Fig. 1A, but with no mismatch), all labelled at the 5' termini of their lower strands (as designated in Fig. 1). Following reaction with extract, 4 of the 45-bp DNA substrates (Figs. 1A, 1C, 1D, and 1E) that contained G:T mispairs, were incised and gave rise to smaller DNA fragments (lanes 2, 5, 6 and 7). One, substrate 1B (lane 4), also containing a G:T mispair, did not. The migration of the incision fragment in lane 2 was the same as that of the 22 nt *AccI* fragment (lane 1), implying that the phosphodiester bond 5' to the mismatched T had been incised, consistent with the previous reports for a 90 bp substrate (3,4) containing the Fig. 1A sequence. By comparing the migration distances of the incision fragments in lanes 5, 6 and 7 with those of the restriction-digestion markers in lane 3, the same conclusion was reached: in these DNAs (Figs. 1C, 1D and 1E), the incision was also 5' to the mismatched T. By contrast, neither non-mismatched control G:C DNA (lane 8) nor the DNA (Fig. 1B), containing a G:T base mismatch at a GpG dinucleotide, (lane 4) were incised by the activity. 5-methylcytosine 5' to the G involved in the G:T mispair did not affect the incision by the mismatch activity (compare lanes 6 and 7). The results are consistent with previous findings (3–6)

that the G:T base mismatch-specific activity in extracts of human cells incises the phosphodiester bonds immediately 5' and 3' to the mismatched T. What is new is the lack of such activity toward the substrate in Fig. 1B in which the G:T mismatch involves the 3' G of a GpG dinucleotide.

The experiment in Fig. 2 did not control for two possibilities, i) that the substrate in Fig. 1B might be improperly constructed, or ii) that, by itself, the Fig. 1B substrate might block incision activity either by virtue of its own peculiar structure, or by contaminating factors. We again prepared double stranded substrate 1B, labelled at the 5' terminus of the bottom strand,

Figure 1. 45-bp DNA substrates. (A) Sequence from the polylinker region of M13mp9 (base pairs 6158–6202). This sequence is contained within the 90 bp substrate used by Wiebauer and Jiricny (3,4). A single G:T base mismatch, located at position 23 of the top strand, modifies the 4th base within the SalI and AccI restriction recognition sites. Incision sites for a DNA in which the mismatch has been repaired to a G:C pair are shown by arrows. (B) Sequence based on H-ras oncogene and first used by Voigt et al. (21). A single G:T mispair at position 20 from the 5' terminus of the top strand replaces the first base pair in a HhaI recognition sequence; restriction incision sites for DNA in which G:T is converted to G:C are shown by arrows. (C) The top strand is the same as in B. The bottom strand is altered such that a single G:T base mismatch exists within a NarI restriction site at position 22 from the 5' end. The digestion sites for the same DNA in which G:T mispairs are converted to G:C pairs are shown by arrows. (D) The top strand is the same as in B. The bottom strand is altered such that a single G:T mismatch at position 25 of the top strand exists within a HpaII site. The digestion sites for the corresponding G:C-containing DNA are shown by arrows. (E) The sequences is the same as in D except that the C 5' to the G involved in G:T mispair is methylated in its 5 position. (F&G) The sequences are same as B, but do not contain mispairs. In F the internal C's in the HpaII incision sites are methylated. Fig. 1G represents DNA without methylation. Numbers near the restriction enzyme abbreviations indicate length in nt of fragments expected after digestion of repaired substrates that are 5'-labelled in their bottom strands.

to be certain it had been properly prepared. This new substrate failed to serve as a substrate for incision. The second possibility was ruled out by the following competition experiment.

Competitive inhibition of G:T base mismatch specific incision activity by 45-bp DNAs

We investigated the ability of selected G:T mismatch-containing DNAs, including the substrate in Fig. 1B, to compete for G:T mismatch-specific incision activity. To do this, we analyzed incision by A1235 extract of the DNA substrate in Fig. 1A (5' terminally labelled in the strand containing the mismatched T) containing a G:T base mismatch, in the absence and presence of increasing amounts of either unlabelled control G:C DNA (Fig. 1G) or mismatched DNAs (Fig. 1B-F). The result is shown in Fig. 3. In the absence of competitive DNA substrate (lane 2), mismatch incision of 45-bp DNA substrate in Fig. 1A generated a 22 nt fragment as in Fig. 2. Addition to reaction mixtures of any of the three 45 bp DNAs that contain G:T mismatches at CpG sites (Fig. 1C, D and E) resulted in less than control amounts of product, showing that these substrates competed well

Figure 2. Incision of 45-bp DNA substrates containing single G:T mispairs. Each duplex DNA was prepared with a 5'-labelled bottom strand (labelled strands are designated *) by hybridization with an unlabelled top strand (G strand). Mixtures of 2 ng of labelled 45-bp DNAs and 6 ng of unlabelled 45-bp DNAs (of exactly the same sequences) were incubated with 10 µg cell-free extract from A1235 in 50 µl incision buffer for 12 hours at 30°C. The DNA incision products were analyzed as described in Materials and Methods by electrophoresis through a 12% sequencing gel along with markers. The markers in lane 1 were prepared by digesting a DNA similar to that in Fig. 1A (but with a G:C replacing the G:T mispair) with SalI and AccI. The markers in lane 3 were produced by digesting the DNA in Fig. 1G with HpaII, NarI and HhaI. The intensities of the incision signals in lanes 6 and 7 are less than those in lanes 2 and 5 because of the different amounts of DNAs loaded onto the gel. The capital letters in parentheses refer to the substrate structures in Figure 1. (That the 22 nt fragment in lane 1 and the 23 nt fragment in lane 3 show approximately the same migration distance is likely due to differences between the structures of the DNAs.) The vertical line separates lanes in which M13mp9-based substrates were analysed from lanes in which H-ras-based substrates were analysed.

Figure 3. Inhibition of G:T base mismatch incision activity by competition. Incision assays were performed using as substrate the 45-bp DNA (shown in Figure 1A) containing a single G:T base mismatch and labelled in 5' end of its bottom strand. Reaction conditions were as for Figure 2 except that the mixture of 2 ng labelled DNA plus 6 ng unlabelled DNA was replaced by 2 ng of labelled DNA plus 40 ng of 1 kb-DNA ladder (BRL). In lanes 3–20, incision assays were carried out in the presence of increasing amounts of unlabelled 45-bp DNA substrates. The capital letters in parentheses refer to the substrate structures in Figure 1.

for the incision activity (Fig. 3, lanes 9–14 and 18–20). (One of these DNAs, that depicted in Fig. 1C, appeared to compete less well for the activity than did the other two competing DNAs; compare lanes 9–11 with lanes 12–14 and 18–20). By contrast, 5 and 10 ng of unlabelled control 45 bp DNAs (lacking mismatches as in Fig. 1G or fully methylated as in Fig. 1F) produced little change in the amounts of reaction product (compare lane 2 with lanes 3 and 4, or lanes 15 and 16). With 20 ng of control competitor DNA (10-fold the amount of labelled substrate) the amount of product was somewhat reduced with the non-methylated G:C competitor (lane 5), but not affected by the CpG-methylated DNA (Fig. 1F, lane 17). The amounts of reaction product in the presence of the Fig. 1B competitor DNA bearing a G:T mismatch (lanes 6–8) were indistinguishable from those obtained when adding non-mismatched G:C DNA as competitor (Fig. 1G). The result reveals that the DNA having a G:T mismatch at a GpG site (Fig. 1B) served neither as a substrate for G:T mismatch incision activity nor blocked G:T mismatch incision.

Repair of G:T mispairs to G:C pairs in 45-bp DNAs

These experiments, whose results are presented in Figs. 2 and 3, assay only incision in the strand containing the mismatched T and do not address the question of whether complete restoration of a G:C product would occur in an environment favoring DNA polymerase and DNA ligase action. Human cell-free extracts, in the presence of Mg^{++} , ATP and the 4 DNA-dNTPs, can perform complete repair by converting a G:T mispaired substrate to a ligated G:C-containing product (3,4). The assay is based on the restoration, by conversion of G:T mispairs to G:C pairs, of restriction sites whose digestion is blocked by G:T mismatches. The G:T mismatched substrates (Fig. 1A–E) were used in the assay along with cell-free extract from A1235 cells. The result is shown in Fig. 4. Without treatment by the extract, the G:T-mispaired DNA substrate (Fig. 1A) yielded no restriction fragment after Sall digestion (lane 1) because the G:T mispair within the Sall site blocked digestion. After the reacting with

Figure 4. Repair of single G:T mispair by A1235 extracts. 45-bp DNAs (Fig. 1A–E), labelled in the 5' ends of their bottom strands (2 ng plus 6 ng of unlabelled DNA of exactly the same structure), were incubated with 10 μ g A1235 cell-free extract in repair synthesis buffer in a total of 50 μ l. Repair synthesis was assayed as described in Materials and Methods. Lanes 3, 6 and 10 show restriction fragments generated by Sall, HpaII and NarI digestion of repaired DNAs. The capital letters in parentheses refer to the substrate structures in Figure 1. As in Figure 2, the fact that restriction fragments from substrate A migrate differently from fragments of the same length from the other substrates is due to differences in base composition.

extract, a fraction of the product (lane 2) was digested by Sall into smaller fragments having the same electrophoretic behavior as the 5' terminally labelled 21-nt restriction fragment from bottom strand (lane 3). The 22-nt and 23-nt minor products, also present without Sall digestion (lane 2), and less intense than the 21-nt band, probably represent (i) substrate incised by the G:T

Figure 5. DNA substrates with G:T mismatches involving the 3' G of GpG, CpG, ApG and TpG dinucleotides (NpG/N'pT substrates). A–D, 45-mer DNAs with structure based on the H-ras oncogene sequence (21). Boxes enclose the G:T mismatch and the base pair defined by the nucleotide 5' to mismatched G.

Figure 6. Competition for mismatch incision by unlabelled CpG:T DNA (Fig. 1A) with labelled substrates of the type NpG/N'pT. Samples prepared with 2 ng of the four substrates in Fig. 5 (³²P-terminally labelled in their bottom strands) mixed with 0, 5, and 10 ng of unlabelled M13mp9 polylinker substrate bearing a G:T pair (Fig. 1A), were incubated for 12 hrs with 10 μg A1235 extract as described in Fig. 2 and analyzed as in Materials and Methods. '20 mer' indicates the position of the HpaII fragment of non-mismatched control DNA, Fig. 1G, labelled in the 5' terminus of the bottom strand.

mismatch activity, but not further repaired (22-nt band) and (ii) incised substrate to which 1 nucleotide was added, but upon which DNA ligase did not act (23-nt band). Similarly, the DNA shown in Fig. 1D, which has a sequence quite different from the DNA in Fig. 1A except for the G:T base mismatch, is resistant to HpaII digestion (lane 4). Only after incubation with extract was it digested to produce the 20-nt 5' labelled HpaII bottom strand fragment (lane 6). As previously, the minor fragments in lane 5, i.e. the 20 and 21-nt fragments likely represent lesser fractions of incised DNA, and incised, but unligated, DNA to which one

Figure 7. Competition for incision by unlabelled H-ras based DNAs (Fig. 5), of the type NpG/N'pT, with labelled M13mp9-based substrate (Fig. 1A) having a CpG:T mispair. Samples with 2 ng of substrate (Fig. 1A; ³²P-labelled in the 5'-terminus of the bottom strand) mixed with 0, 5 and 10 ng of unlabelled DNAs (Fig. 5A-D) were incubated with 10 μg A1235 cell-free extract protein as described in Fig. 2, and analyzed as in Materials and Methods. '22 mer' indicates the migration distance of the AccI digestion product of non-mismatched control substrate (Fig. 1A).

nucleotide had been added. In lanes 7 and 8, the DNA is the same as in Fig. 1D except that the C just 5' to the G involved in the mispair with T is 5' methylated, thereby blocking cleavage by HpaII. However, smaller fragments (of 20- and 21-nt size and intensity similar to the fragments in lane 5) indicate that incision and, presumably, nucleotide incorporation events likely took place. (We did not digest this DNA with MspI). As seen in lanes 9 and 10, the DNA of Fig. 1C, that also contains a G:T mismatch in a CpG sequence, is cleaved by the NarI restriction endonuclease only after treatment by extract. Thus we conclude that repair of the G:T mismatch occurred in this substrate. Lane 10 shows that the amount of Fig. 1C DNA that is repaired is less than of the amounts of Fig. 1A or D DNAs that are repaired (lanes 3 and 6), paralleling the relatively reduced ability of Fig. 1C substrate to compete for G:T mismatch activity (compare Fig. 3 lane 9 with lanes 12 and 18). By contrast to the repair of these DNAs, the DNA with a G 5' to the G involved in mismatch (Fig. 1B) was not cleaved at all by HhaI digestion either before or after treatment with extract (lane 11 and 12). A control experiment showed that HhaI did digest the non-mismatched Fig. 1G DNA (data not shown). Thus the substrate in Fig. 1B, containing the G:T mismatch at a GpG site, failed to serve as a substrate either for G:T mismatch incision or complete repair, and failed to compete for G:T mismatch incision activity.

Mismatch incision of a G:T at one site of the top strand is independent of the base pair 5' to the mismatched G

To determine whether the lack of G:T mismatch repair of the Fig. 1B substrate is due to possible specificity of the human G:T mismatch repair system for substrates in which a C is 5' to the mismatched G, we constructed three additional H-ras-based substrates of the type NpG/N'pT, and bearing a mismatch at position 25 of the top strand. These are depicted in Figs. 5B–5D. They contain an A, T or G 5' to the mismatched G. The substrate in Fig. 1D with a C 5' to the mismatched G is shown again as Fig. 5A for comparison. The Fig. 5A–D substrates were ³²P-

labelled in the 5' termini of their bottom strands and incubated with A1235 extract. In an initial experiment (data not shown) all 4 substrates were incised. To assure that such incision was not being effected by different mismatch repair systems having specificities dependent upon the base 5' to the mismatched G, unlabelled DNA bearing a CpG:T mismatch (Fig. 1A) was added at 0, 5 and 10 ng to compete with 2 ng of the labelled substrates in incision reactions. The results of the experiment are shown in Fig. 6. The somewhat variable yields of incised product in the absence of competing DNA (similar to the result of the initial experiment) could be caused by unequal labelling efficiency of the 5'-termini of the substrates. With 5 and 10 ng of unlabelled DNAs competition was strong and product yields were, within experimental error, reduced to control levels. Such competition was independent of which nucleotide is 5' to the mismatched G in the labelled substrates, and excludes the possibility of different mismatch repair systems alluded to earlier.

To determine whether any of the 4 NpG/N'pT DNAs in Fig. 5A–D may be better substrates than others, competition experiments involving the same labelled substrate (to ensure quantifiability) were done. The M13mp9-based substrate (Fig. 1A) was 5'-labelled on the bottom strand and 2 ng was used as substrate for reactions to which 5 or 10 ng of the unlabelled NpG/N'pT substrates were added as competitor DNAs. The results are shown in Fig. 7. Control samples in lanes 1 and 2 contained no competing DNA: the sample in lane 2 was treated with extract, while that in lane 1 was not. The addition of 5 ng of unlabelled ApG:T or GpG:T competing DNAs to the reaction (lanes 3 and 5) caused little diminution in the amount of incised product relative to the extract-treated control (lane 2), whereas the TpG:T and CpG:T mismatched DNAs were more effective at the same concentration (lanes 7 and 9). The addition of 10 ng of any competing DNA to the reactions markedly suppressed product formation (lanes 4, 6, 8 and 10). The experiment was performed twice with nearly identical results. Accordingly, while the base pair 5' to the mismatched G may not govern the specificity of G:T mismatch repair absolutely, it may do so relatively since it appears to influence the amount of incised product.

DISCUSSION

We have presented evidence that the G:T mismatch repair activity in human cellular extracts prepared by the method of Manley et al (19) and assayed as previously (6) can incise and repair DNA substrates having G:T mispairs in a DNA site-specific manner. The results in Figs. 2–4 show that substrates with G:T mispairs at several sites are substrates for such incision and repair, but that a substrate with a G:T at another site (Fig. 1B) is not. The conclusion that the Fig. 1B DNA is not a substrate for G:T mismatch repair in our experiments extracts is supported by three results: (i) the activity in the cell-free system failed to incise the Fig. 1B DNA substrate (Fig. 2), (ii) mismatch incision activity was not competitively inhibited by the Fig. 1B substrate (Fig. 3), and (iii) restoration of G:T mispairs to G:C pairs did not occur on the Fig. 1B substrate (Fig. 4). The competition experiments of Fig. 3 showed that inability to incise substrate in Fig. 1B was not caused by factors in the substrate that inactivated the enzyme. Further, a repeat preparation of the Fig. 1B DNA again failed to serve as substrate for the incision reaction. The results shown in Figs. 6 and 7 show that when human extract performed incision at a G:T mismatch at one of the digestible sites, there was

evidence for relative, but not absolute, dependence upon the base pair 5' to the mismatched G. Thus failure to repair the G:T mismatch in the Fig. 1B substrate is not due to any specificity that the human G:T mismatch repair system has for mismatches occurring at CpG sites. We do not now understand why the DNA in Fig. 1B is refractory to human G:T mismatch repair activity, but possibilities include strong binding by a DNA context-specific protein (or proteins) present in the cell-free extract.

The substrate in Fig. 1C was incised well as shown in Fig. 2, competed relatively poorly for incision as shown in Fig. 3 and was repaired relatively poorly as shown in Fig. 4. When different labelled substrates are prepared, variations in substrate specific activity achieved by T4-kinase labelling, for example, might cause variations in the apparent amounts of incision product in the experiment of Fig. 2. The competition experiment is the best for such comparative purposes because the same labelled substrate is used throughout, and the concentrations of the unlabelled competing DNAs are known with good accuracy. Thus it is possible that the substrate in Fig. 1C is actually a poorer substrate for mismatch repair than the substrates in Figs. 1A and 1D, but further work is required to make this conclusion.

Taken in conjunction with the known specificity of cytosine 5-methylation at CpG sites, our findings constitute evidence that human systems do not both 5-methylate cytosines and repair G:T mismatches at similar DNA sequences as is the case in *E. coli* K. In human cells, spontaneous deamination of 5-methylcytosine can produce G:T mismatches only from m⁵CpG dinucleotides, so the relatively broad specificity of the human activity may reflect an evolutionary selection for G:T substrate specificity by potentially mutagenic reactions other than m⁵C deamination that also give rise to G:T mispairs, including errors of DNA replication or recombination intermediates. That CpG is not the sole site of G:T mismatch repair in human cells suggests that the human G:T mismatch repair pathway may not be functionally analogous to the *E. coli* K very short patch repair pathway (12–15) in other respects, and this is likely so. m⁵C methylation of CpG sites is believed to control transcription in human cells (16,23), while the function (other than targeting methylation and mismatch repair) of methylated 5'-CC(A/T)GG-3' sites in *E. coli* K strains is unknown, but almost certainly would be known if it were transcriptional regulation.

Wiebauer and Jiricny published evidence that human G:T mismatch repair activity involves a mismatch-specific thymine glycosylase and AP endonuclease activity (4). Glycosylases such as uracil DNA glycosylase and hypoxanthine glycosylase recognize substrates that are produced by deamination of either cytosine or adenine bases in DNA. That the human G:T base mismatch activity releases a normal base, thymine, is surprising (3). Non-mismatched DNA with a fully methylated HpaII site (Fig. 1F) was not a substrate for *in vitro* mismatch repair (as is the case with the *E. coli* K Vsr protein; ref. 15), while the hemimethylated analogue with a G:T mismatch was a substrate. These results show that the glycosylase does not simply remove any 5-methylated base opposite guanine, but rather shows some specificity for thymine, possibly related to the thymine 4-oxygen and/or N-3 proton, both lacking in cytosine. Earlier data showing that both G:T and m⁶G:T pairs are substrates for G:T mismatch repair gave rise to the suggestion that G:T mismatch recognition also involves the 2-amino group of the purine involved in the mismatch (6). From these considerations we envision a scenario in which a protein (or proteins) simultaneously and sequentially scan base pairs both in the minor groove for purines with 2-amino

groups and in the major groove for pyrimidines with 4-oxygens (and or 3-protons). Such recognition of G:T mismatches may not be a function of the glycosylase but rather of a separate mismatch binding protein, possibly similar to the *E.coli* MutS mismatch binding protein (24). Several such human base mismatch-specific binding proteins have been reported (25,26).

ACKNOWLEDGEMENTS

This study was supported by National Cancer Institute Award CA49936, a National Cancer Institute of Canada Award, and an Alberta Heritage Foundation for Medical Research Scholarship/Establishment grant to R.S.D.

REFERENCES

1. Shapiro, R. and Klein, R.S. (1966) *Biochemistry* **5**, 2358–2362.
2. Duncan, B.K., and Miller, J.H. (1980) *Nature* **287**, 560–561.
3. Wiebauer, K. and Jiricny, J. (1989) *Nature* **339**, 234–236.
4. Wiebauer, K. and Jiricny, J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 5842–5845.
5. Yeh, Y.-C., Chang, D.-Y., Masin, J., and Lu, A.-L. (1991) *J. Biol. Chem.* **266**, 6480–6484.
6. Sibghat-Ullah and Day, R.S., III. (1992) *Biochemistry* **31**, 7998–8008.
7. Brown, T.C. and Jiricny, J. (1987) *Cell* **50**, 945–950.
8. Price, A. and Lindahl, T. (1991) *Biochemistry* **30**, 8631–8637.
9. Chen, D.S., Herman, T., and Demple, B. (1991) *Nucleic Acids Res.* **19**, 5907–5914.
10. Leonard, G.A., Thompson, J., Watson, W.P. and Brown, T. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 9573–9576.
11. Karran, P. and Bignami, M. (1992) *Nucleic Acids Res.* **20**, 2933–2940.
12. May, M.S. and Hattman, S. (1975) *J. Bacteriol.* **122**, 129–138.
13. Lieb, M. (1985) *Mol. Gen. Genet.* **199**, 465–470.
14. Sohail, A., Lieb, M., Dar, M., and Bhagwat, A.S. (1990) *J. Bacteriol.* **172**, 4214–4221.
15. Hennecke, F., Kolmar, H., Brundl, K., and Fritz, H.-J. (1991) *Nature* **253**, 776–778.
16. Jones, P.A. and Buckley, J.D. (1990) In G.F. Vande Woude and G. Klein (eds.) *Advances in Cancer Research*, Vol 54. Academic Press, New York. pp 1–23.
17. Holmes, J. Jr., Clark, S., and Modrich, P. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 5837–5841.
18. Thomas, D.C., Roberts, J.D., and Kunkel, T.A. (1991) *J. Biol. Chem.*, **266**, 3744–3751.
19. Manley, J.L., Fire, A., Cano, A., Sharp, P.A., & Geftter, M.L. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 3855–3859.
20. Sibghat-Ullah, Husain, I., Carlton, W., & Sancar, A. (1989) *Nucleic Acids Res.*, **17**, 4471–4484.
21. Voigt, J.M., Van Houten, B., Sancar, A., & Topal, M.D. (1989) *J. Biol. Chem.*, **264**, 5172–5176.
22. Lieb, M. (1991) *Genetics*, **128**, 23–27.
23. Bird, A. (1992) *Cell*, **70**, 5–8.
24. Modrich, P. (1991) *Annu. Rev. Genet.*, **29**, 229–253.
25. Jiricny, J., Hughes, M., Corman, N., and Rudkin, B.B. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8860–8864.
26. Stephenson, C. and Karran, P. (1989) *J. Biol. Chem.* **264**, 21177–21182.