Incision at DNA G·T Mispairs by Extracts of Mammalian Cells Occurs Preferentially at Cytosine Methylation Sites and Is Not Targeted by a Separate G·T Binding Reaction

Shaun Griffin and Peter Karran*

Imperial Cancer Research Fund, Clare Hall Laboratories, South Mimms, Hertfordshire EN6 3LD, UK

Received August 2, 1993; Revised Manuscript Received September 7, 1993®

ABSTRACT: We have investigated the specificities of G·T mismatch binding proteins and of G·T mismatch cleavage in extracts of mammalian cells. G·T mismatch-specific protein:DNA complex formation by cell extracts was independent of the local sequence context of the mismatch. Cell extracts performed similar levels of protein binding to DNA substrates in which a single G·T mispair was preceded by T, G, A, C, or 5-meC. In contrast, incision by extracts of the T-containing strand of a G·T mismatch exhibited a strong sequence specificity and efficient strand cleavage was only observed when the mismatched G was in a CpG sequence. Thus, oligonucleotides containing either $^{CpG}_{GpT}$ or $^{5meCpG}_{GpT}$ were efficiently incised, but not those containing $^{GpG}_{CpT}$, $^{ApG}_{TpT}$, or $^{TpG}_{ApT}$ sequences. Cell lines made resistant to the alkylating agent *N*-methyl-*N*-nitrosourea have previously been found to be defective in a G·T mismatch binding reaction. The defect in binding by extracts prepared from these cells extended to G·T mismatches in several sequence contexts. The variant extracts nevertheless incised G·T mismatches normally suggesting that this particular binding activity is not required for incision. The data indicate that incision by this activity is targeted to the CpG sequences in which G·T mismatches are formed by the mutagenic deamination of DNA 5-methylcytosine. In this regard the repair pathway resembles the very short patch (vsp) repair pathway in *Escherichia coli*.

DNA mismatch repair increases the fidelity of DNA replication (Radman & Wagner, 1986), promotes gene conversion during recombination (Holliday, 1964), and removes inappropriate bases from resting DNA (Jones et al., 1987). *Escherichia coli* employ more than one strategy to correct G·T mismatches in DNA. A general mismatch repair system corrects G·T and other single base mispairs that arise by misincorporation during replication and escape proofreading by DNA polymerases. This pathway is dependent on the *mutH*, *mutS*, *mutL*, and *dam* genes, and the appropriate DNA strand is targeted for correction by a transient undermethylation of A in GATC sequences at the replication fork (for review see Modrich, 1989). An additional correction mechanism, very short patch (vsp)[1] repair, is G·T mismatch-specific. Vsp correction is dependent on the *vsr* gene product and is targeted to DNA sequences wherein thymine bases arise *via* deamination of 5-meC (Lieb, 1991; Hennecke et al., 1991). There is a partial overlap between the two pathways since the *E. coli* MutS and MutL proteins are absolutely required for methyl-directed mismatch repair and, although not essential, enhance the efficiency of vsp repair (Jones et al., 1987; Lieb, 1987). MutS is a mismatch recognition protein that binds specifically to all single base mispairs in DNA and through protein–protein interactions coordinates the enzymes involved in excision of the mismatched base (Modrich, 1989). The MutY protein of *E. coli* also recognizes some single base mispairs, in particular A·G, but genetic and biochemical evidence indicates that this protein acts specifically to remove adenine misincorporated opposite 8-oxoguanine in DNA (Michaels et al., 1992).

Mammalian cells can repair all DNA mismatches in transfected heteroduplex DNA although the efficiencies of correction of individual mispairs vary widely (Brown & Jiricny, 1988). G·T mismatches are the most efficiently rectified and more than 90% of progeny DNA molecules exhibit unidirectional repair of G·T to G:C base pairs (Brown & Jiricny, 1987). Extracts of human cells contain a thymine-specific DNA glycosylase that removes the mispaired thymine from a G·T mismatch (Wiebauer & Jiricny, 1990). This appears to be the initial step of G·T-specific, short patch excision repair analogous to the vsp pathway of *E. coli* and is assayed in cell extracts as specific cleavage of duplex oligonucleotides containing single G·T mispairs (Wiebauer & Jiricny, 1989; Sibghat-Ullah & Day, 1993). Repair is completed by the removal of the apyrimidinic site followed by replacement of the single missing nucleotide and ligation. Other mismatches are repaired by a more general pathway that involves a long excision tract (Holmes et al., 1990; Varlet et al., 1990; Thomas et al., 1991). An A·G-specific incision activity has also been demonstrated in partially purified extracts of human cells (Yeh et al., 1991). This counterpart of the *E. coli* MutY protein may be involved in a separate mismatch correction pathway for oxidatively damaged DNA. Human cell extracts also contain proteins that bind specifically to DNA mismatches. One protein–DNA complex is formed preferentially with DNA containing G·T mispairs (Jiricny et al., 1988). A distinct activity selectively binds to DNA with single A·C, T·C, or T·T mispairs (Stephenson & Karran, 1989). Together, these functions are similar to those of the *E. coli* MutS protein, but the role of mismatch recognition proteins in either short patch or general mismatch repair in mammalian cells is not known.

Some mammalian cell lines that have acquired resistance to methylating agents are selectively defective in G·T mismatch binding assayed in cell extracts. The cell lines exhibit a spontaneous mutator phenotype, and this suggests that the G·T binding activity might act in a mismatch correction

® Abstract published in *Advance ACS Abstracts*, November 15, 1993.

[1] Abbreviations: EDTA, ethylenediaminetetraacetate; DTT, dithiothreitol; Pipes, piperazine-*N*,*N*′-bis[2-ethanesulfonic acid]; 5-meC, 5-methylcytosine; vsp, very short patch repair.

0006-2960/93/0432-13032$04.00/0

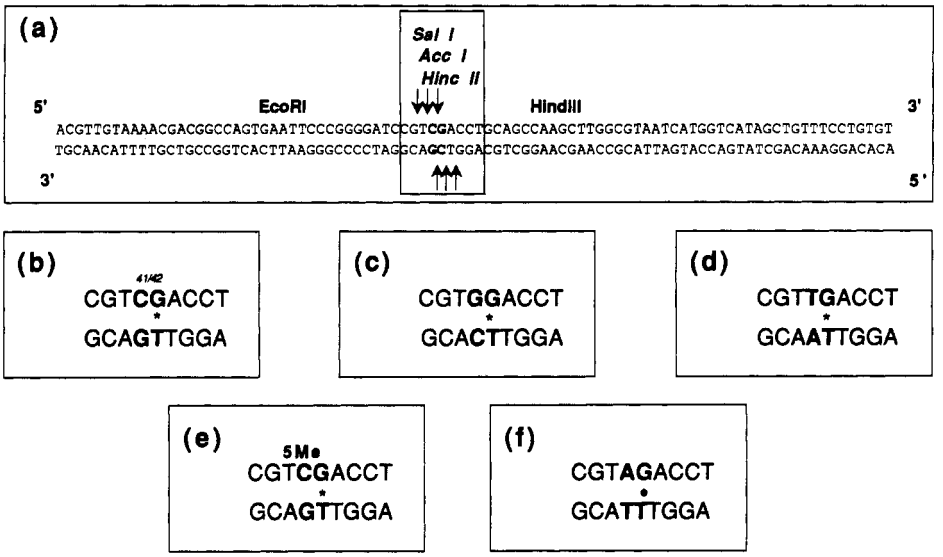

FIGURE 1: Sequences of 90mer duplex oligonucleotides used in this study. Both strands of the perfectly paired duplex are shown in (a). The mismatch site (position 42 of the upper strand as written) was replaced by a G·T mismatch preceded on the 5′ side by the base pairs shown (b–f).

pathway (Branch et al., 1993). The association of methylation tolerance with a mutator phenotype has recently been confirmed in the human cell line MT1 (Kat et al., 1993). The identical phenotype of the cells deficient in G·T binding and MT1 cells that are unable to correct all single base mispairs *in vitro* provides further evidence that the G·T binding activity is involved in mismatch repair. We have used extracts from the G·T binding-deficient variants to investigate whether this activity is required to initiate correction by targeting the mismatch incision activity to G·T mismatches. We show here that incision at G·T mismatches occurs normally in extracts of the G·T binding-deficient variants and that, whereas formation of the G·T mismatch:protein complex is essentially independent of the sequence context of the mismatch, G·T mismatches are efficiently incised only when the mismatch is located in CpG or 5-meCpG dinucleotides in DNA. The data are consistent with a specific pathway for repair of G·T mismatches that is targeted by the immediate sequence context of the G·T mispair and not by G·T mismatch binding protein(s).

MATERIALS AND METHODS

Cells and Cell Culture. HeLaMR, CHOMT$^+$, and Clone B cells (Aquilina et al., 1988, kindly provided by Dr. G. Aquilina, Istituto Superiore di Sanitá, Rome) were maintained in suspension culture in Dulbecco's modified Eagles medium containing 10% calf serum. The TK$^-$ variant of the Burkitt's lymphoma line Raji and the *N*-methyl-*N*-nitrosourea-resistant line RajiF12 (Branch et al., 1993) were maintained in RPMI 1640 containing 10% calf serum.

Oligonucleotide Synthesis. 90mer oligonucleotides (Figure 1) (Wiebauer & Jiricny, 1989) were synthesized on an Applied Biosystems 380B DNA synthesizer. The 5meC phosphoramidite was obtained from Cruachem, Glasgow, UK. All oligonucleotides were deprotected under mild conditions by heating in 30% ammonia solution at 55 °C for 8 h and detritylated in acetic acid. Each oligonucleotide was purified by reverse-phase HPLC followed by elution from 10% denaturing polyacrylamide gels. Bottom strands (as written in Figure 1) were 3′-end-labeled by addition of a single [^{32}P] dideoxyAMP residue using the Amersham 3′-end labeling kit. Top strands were 5′-end-labeled by T4 polynucleotide kinase with [γ-^{32}P]ATP. Duplex substrates were prepared by annealing a 5-fold molar excess of the unlabeled strand to the appropriate labeled strand in 10 mM Tris·HCl pH 8.0, 1 mM EDTA, 0.25 M NaCl. Oligonucleotides were heated to 100 °C and cooled slowly to room temperature.

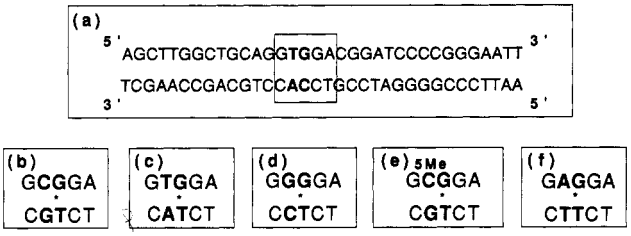

FIGURE 2: Sequences of 34mer duplex oligonucleotides used in this study. Both strands of the perfectly paired duplex are shown in (a). The mismatch site (position 16 of the upper strand as written) was replaced by a G·T mismatch preceded on the 5′ side by the base pairs shown (b–f).

34mer oligucleotides (Jiricny et al., 1988; Stephenson & Karran, 1989; Figure 2) were synthesized in similar fashion and used without further purification. For mismatch binding or cleavage assays the appropriate single strands were labeled at their 5′ ends by T4 polynucleotide kinase with [γ-^{32}P]ATP. Annealing reactions were carried out as above.

Cell Extracts and Assays. G·T mismatch nicking assays were carried out using replication extracts prepared by a modification of published procedures (Li & Kelly, 1985). Extracts for mismatch binding assays were prepared as described (Stephenson & Karran, 1989). Extracts were stored as aliquots at −80 °C and used before a maximum of two freeze–thaw cycles.

Optimal assay conditions for G·T nicking activity by cell extracts were determined empirically. In the standard assay, 40 fmol end-labeled duplex oligonucleotide was incubated with cell extract (0–10 μg) in 50 mM Pipes·NaOH pH 6.7, 10 μM $ZnCl_2$, 0.5 mM EDTA, 0.5 mM DTT in a total volume of 20 μL. Incubations were generally carried out for 16 h at 30 °C and reactions were stopped by the addition of 20 μL formamide containing electrophoresis marker dyes. Samples were denatured by boiling and 10 μL analyzed on 12% denaturing polyacrylamide gels. Markers were derived by restriction digestion of the control oligonucleotide (Figure 1a) 3′-end-

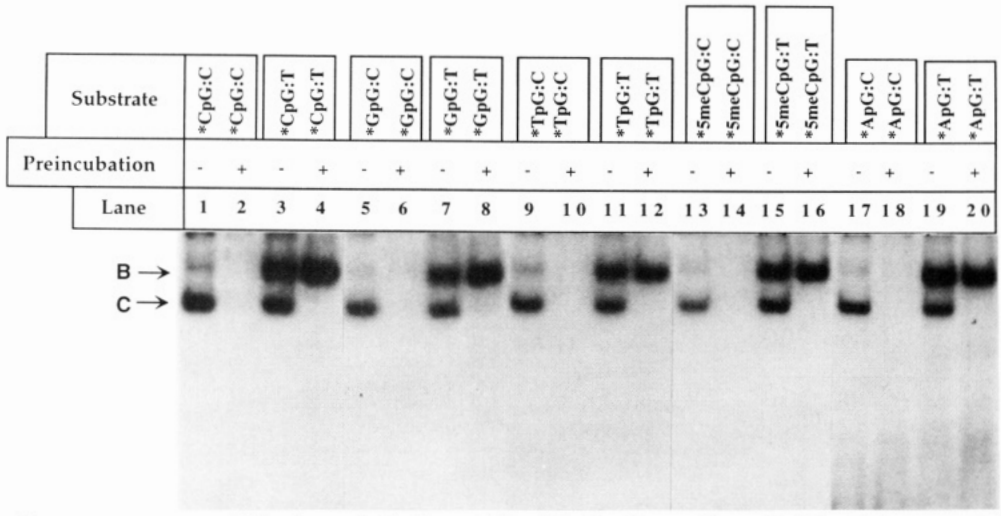

FIGURE 3: Effect of sequence context on G·T mismatch binding activity. Duplex 34mer oligonucleotides that were perfectly matched with a G:C pair (lanes 1, 2, 5, 6, 9, 10, 13, 14, 17, and 18) or with a single G·T mismatch at position 16 (lanes 3, 4, 7, 8, 11, 12, 15, 16, 19, and 20) in which the mismatched G was immediately preceded by a C (lanes 3 and 4), a G (lanes 7 and 8), a T (lanes 11 and 12), a 5-meC (lanes 15 and 16), or an A (lanes 19 and 20) were incubated with 15 μg RajiTK$^-$ cell extract for 20 min at 20 °C. Binding to the oligonucleotides was analyzed by nondenaturing polyacrylamide gel electrophoresis as described in Materials amd Methods. Samples designated (+) were preincubated for 5 min at 20 °C with a 5-fold excess of nonradioactive matched competitor (CpG:C) oligonucleotide before addition of the labeled substrate in order to suppress nonspecific binding. (*) denotes the labeled strand. To improve resolution, electrophoresis was carried out until the free oligonucleotide had migrated out of the gel.

labeled in the bottom strand or 5′-end-labeled in the top strand. The gel was exposed to X-ray film. Quantitation was performed by analysis of the autoradiograms using a scanning densitometer.

CHOMT$^+$ and the related Clone B cell extracts contained more nonspecific nuclease activity than human cell extracts and when incubated under the standard reaction conditions, degraded the substrate. This degradation could be largely avoided by performing the assay in 50 mM Pipes·NaOH pH 7.0, 10 μM $ZnCl_2$, 0.5 mM EDTA, 0.5 mM DTT, and 80 mM NaCl. Although some breakdown of the substrate was still seen, the extent was not significantly different between CHOMT$^+$ and Clone B cell extracts and did not hinder identification of the reaction product.

Mismatch binding assays using 34mer oligonucleotides (Figure 2) and preparation of cell extracts were carried out as previously described (Stephenson & Karran, 1989). Briefly, 34mer oligonucleotides were incubated with cell extract for 20 min at 20 °C. Binding to the oligonucleotides was analyzed by 6% nondenaturing polyacrylamide gel electrophoresis. To improve resolution electrophoresis was usually continued until the free oligonucleotide had migrated out of the gel (Stephenson & Karran, 1989). Samples were normally preincubated for 5 min at 20 °C with a 5-fold excess of nonradioactive matched competitor to suppress nonspecific binding. The products were detected by autoradiography.

RESULTS

Sequence Specificity of the G·T Binding Activity. Mammalian cell extracts contain protein(s) that bind to DNA fragments containing G·T mismatches and delay their migration through nondenaturing polyacrylamide gels. The use of substrate or competitor oligonucleotides of different sequences previously suggested no preferred sequence context for G·T mismatch binding (Jiricny et al., 1988; Stephenson & Karran, 1989). We used the oligonucleotide duplexes shown in Figure 2 to examine the sequence dependence in more detail. Extracts of the Burkitt's lymphoma cell line RajiTK$^-$ produced two major delayed migrating complexes when combined with end-labeled oligonucleotides containing a single G·T mismatch. The more rapidly migrating complex (C in Figure 3) was formed with all substrates including perfectly matched controls and could be removed by a prior incubation of the extract with a nonradioactive, perfectly matched oligonucleotide (e.g., Figure 3, lanes 1 and 2). The second complex (B in Figure 3) was formed preferentially with substrates containing a G·T mismatch and was unaffected by the matched competitor oligonucleotide (e.g., Figure 3, lanes 3 and 4). This complex has previously been shown to be removed by inclusion of unlabeled competitor G·T heteroduplex in the incubation reaction (Jiricny et al., 1988). Substrates containing a G·T mismatch in which the mismatched G was 3′ to a C (Figure 3, lanes 3 and 4), a G (lanes 7 and 8) a T (lanes 11 and 12), a 5-meC (lanes 15 and 16), or an A (lanes 19 and 20) were complexed to similar extents by Raji cell extracts. These experiments were carried out at near saturating amounts of extract protein (15 μg). To obtain a more quantitative comparison of binding among the different sequence contexts, we carried out the binding reaction at several different extract concentrations. Figure 4a shows that protein binding to G·T mismatches in either 5-meCpG or GpG sequences exhibits a similar dependence on extract concentration. Binding to G·T mismatches in the other sequence contexts as a function of extract concentration was not significantly different (data not shown).

The human cell line RajiF12 and the hamster line CHO Clone B are derivatives of the RajiTK$^-$ and the CHOMT$^+$ lines, respectively. Extracts from these cells are defective in G·T mismatch binding but bind normally to A·C mismatches. The G·T mismatch binding defect in RajiF12 was observed with G·T mismatches in all of the sequence contexts tested. No binding was observed to oligonucleotides in which the mismatched G was preceded by A, G, T, C, or 5-meC. An example using oligonucleotides with G·T mismatches in either TpG or 5-meCpG sequences is shown in Figure 4b.

The data indicate that proteins in extracts of Raji cells bind to G·T mismatches in either CpG, GpG, ApG, TpG, or 5-meCpG sequences with similar efficiencies and that extracts of the variant line RajiF12 are unable to bind to G·T mismatches in any of the sequence contexts tested.

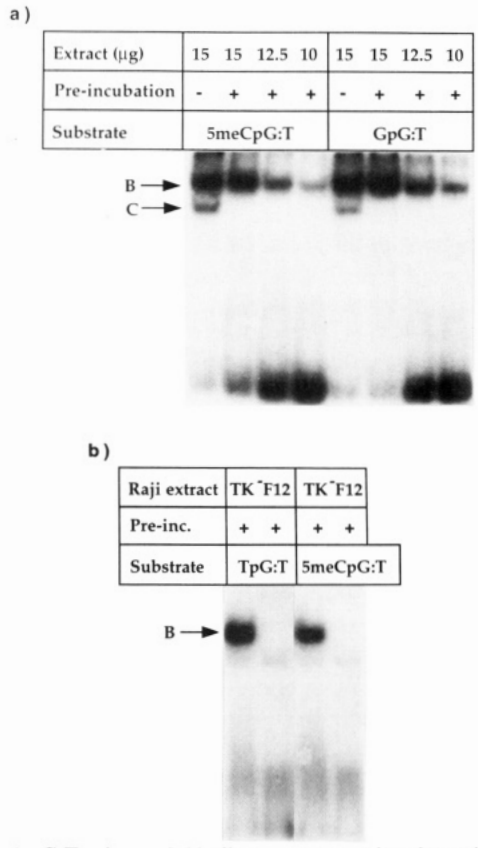

FIGURE 4: G·T mismatch binding: concentration dependence and absence in variant cell lines. (a) Protein concentration dependence. Duplex 34mer oligonucleotides containing either a G·T mismatch in a 5MeCpG or a GpG sequence were incubated with varying amounts of Raji TK⁻ extract as shown. The reaction was carried out as described in Materials and Methods. The most rapidly migrating species is uncomplexed oligonucleotide that was allowed to migrate to the end of the gel. (b) Binding by extracts of variant cells. Duplex 34mer oligonucleotides containing a G·T mismatch in either a TpG or a 5MeCpG sequence were incubated with a 15 μg of Raji TK⁻ or RajiF12 extract as shown. Samples were preincubated for 5 min at 20 °C to remove nonspecific binding before addition of the labelled oligonucleotide. The free oligonucleotide was allowed to migrate out of the gel.

G·T Mismatch Incision by Replication Extracts. Incision of G·T mismatched substrates has previously been demonstrated using nuclear or whole-cell extracts. In our hands, replication extracts prepared by the method of Li and Kelly (1985) gave higher nicking activity and were more reproducible. When incubated with duplex 90mer oligonucleotides either matched or containing a single G·T mismatch in a CpG sequence (Figure 1 (parts a and b, respectively)) replication, extracts prepared from HeLaMR cells also incised the T-containing strand. Because this type of extract had not previously been used to assay G·T mismatch incision, we determined the optimal reaction conditions. The reaction exhibited a sharp pH optimum at pH 6.5–6.9 in Pipes·NaOH buffer, and no cleavage was observed at pH 5.5 or 8.5 (data not shown). The reaction was progressively inhibited by increasing NaCl concentrations up to 80 mM in the standard assay and was completely inhibited by the inclusion of either 10 mM $MnCl_2$ or $MgCl_2$ (data not shown) apparently due to degradation of the substrate by nucleases. Since the extracts contained 100 mM NaCl, the assay was standardized to a final concentration of 5 mM NaCl.

Incision was confined to the T-containing strand of the G·T mismatched oligonculeotide. The extent of nicking was dependent on protein concentration up to 10–20 μg per 20-μL reaction (Figure 5). Two cleavage products that migrated on either side of the 43mer marker were observed. At lower protein concentrations (2–3 μg), the longer cleavage product predominated. Higher protein concentrations produced a second band that was apparently one nucleotide shorter. These data are consistent with incision on both sides of the mismatched T with the initial cleavage being immediately 5′ to the mismatch as suggested by Sibghat-Ullah and Day (1993). The slightly faster migration of the markers is an artefact of loading in slightly different buffer conditions.

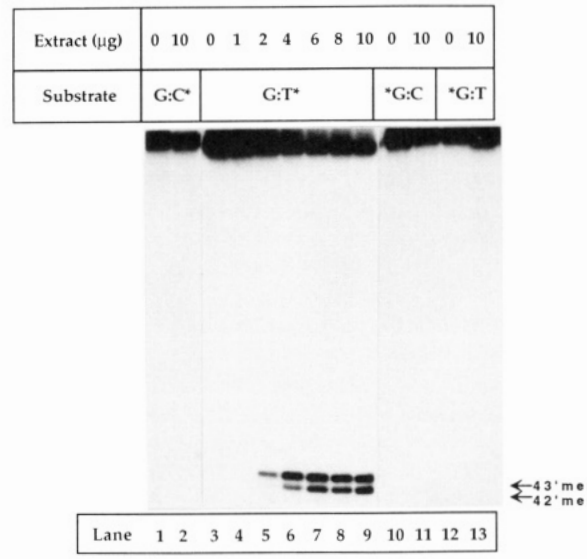

FIGURE 5: Strand specificity of G·T mismatch cleavage by HeLaMR cell extracts. HeLaMR cell extract was incubated at the amounts indicated with matched 90mer oligonucleotide (Figure 1a) 3′-end-labeled in the C-containing strand (lanes 1 and 2). Duplex 90mer containing a G·T mismatch in a CpG sequence (Figure 1b) 3′-end-labeled in the T-containing strand (lanes 3–9). Matched 90mer oligonucleotide 5′-end-labeled in the G-containing strand (lanes 10 and 11). Duplex 90mer containing a G·T mismatch 5′-end-labeled in the G-containing strand (Lanes 12 and 13). Incubation was for 16 h at 30 °C. Oligonucleotides were recovered and analyzed by denaturing polyacrylamide gel electrophoresis. The positions of migration of a 43mer and 42mer oligonucleotide are shown arrowed.

Up to 50% of the substrate was cleaved under optimal assay conditions. Nicking was confined to the T-containing strand and no cleavage of labeled G-containing strands was detected (Figure 5, lanes 11 and 13). The reaction was specific for G·T mismatches, and no incision of either strand of a matched duplex (lanes 2 and 11) or one containing an A·C mispair in which the C-containing strand was 3′-end labeled or the A-containing strand was 5′-end labeled (data not shown) was observed. The reaction was slow, incision was detectable after 1-h incubation at 30 °C but continued to increase for up to 9.3 h (Figure 6). The appearance of the 42mer was always preceded by the 43mer which again suggests that an initial incision 5′ to the mismatched T was followed by a second cleavage on the 3′ side. The strand specificity and the slow kinetics of cleavage are consistent with data obtained with whole cell or nuclear extracts (Wiebauer & Jiricny, 1989; Sibghat-Ullah & Day, 1993) and indicate that replication extracts can initiate incision at G·T mispairs most likely by the previously reported mismatched thymine-DNA glycosylase (Wiebauer & Jiricny, 1990).

To investigate whether the G·T binding activity was required for incision, nicking at G·T mismatches by extracts of G·T binding-deficient variants was compared to their respective parental lines. Extracts of the variant line, RajiF12, exhibited a concentration-dependent nicking activity that was compa-

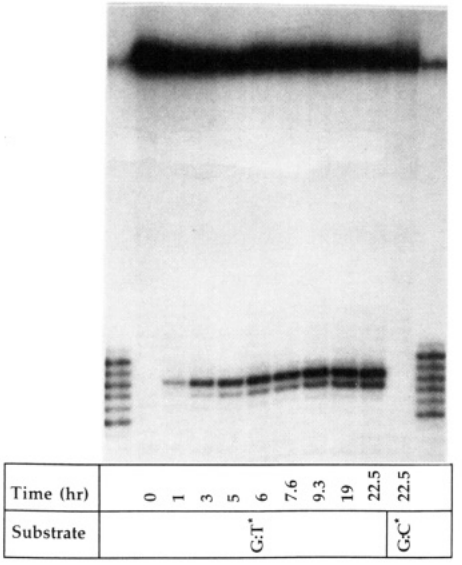

FIGURE 6: Kinetics of G·T mismatch incision by HeLa MR cell extract. 90 μg HeLaMR cell extract was incubated at 30 °C with 360 fmol standard 90mer CpG·T mismatched substrate (Figure 1b) in a total volume of 180 μL. At the times indicated, 20-μL aliquots were withdrawn and processed as described in Materials and Methods. Control CpG:C (40 fmol) oligonucleotide (Figure 1a) was incubated with 10 μg extract as indicated.

rable to the parental RajiTK⁻ line (Figure 7a), and no nicking of the matched control oligonucleotide was observed. Replication extracts prepared from CHOMT⁺ and Clone B contained high levels of nonspecific nuclease activity. This activity could be suppressed to a large extent if the reactions were carried out at pH 7.0 in the presence of 80 mM NaCl. Under these conditions, both CHOMT⁺ and Clone B extracts nicked the T-containing strand of the G·T mismatched heteroduplex. Both extracts incised the substrate to a similar extent (Figure 7b). Only the longer reaction product was seen in this experiment suggesting that the second cleavage reaction proceeded less efficiently under these reaction conditions. Under the same, suboptimal, reaction conditions Raji extracts were also unable to carry out the second cleavage reaction (data not shown). Incision exhibited the same specificity, and no cleavage of the matched control substrate by either parental or variant cell extracts was observed. Thus, incision at G·T mismatches is normal in extracts of two independent variants that are both selectively defective in G·T mismatch binding. These observations indicate that this particular G·T binding activity is unlikely to target the incision enzymes to G·T mismatches.

Effect of Local Sequence Composition on G·T Mismatch Cleavage. To examine the effect of local sequence context on G·T cleavage, we compared the activity of the HeLaMR extract toward a G·T mismatch in which the G was 3′ to a C, a G, a T, a 5meC, or an A in the 90mer (Figure 1b–f). Figure 8 shows that among these four substrates, nicking activity was essentially confined to G·T mispairs in which the G was in a CpG or 5meCpG sequence. Less than 10% of maximum cleavage occurred when the T was opposite the G in a TpG or ApG dinucleotide, and no detectable nicking was observed when the mismatched T was opposite the 3′ G of a GpG dinucleotide.

We investigated whether the apparent preferential cleavage of G·T mismatches in CpG dinucleotides extended to unrelated DNA sequences by assaying nicking of the 34mer oligonucleotides used in the G·T binding experiments. These sequences do not contain regions of extensive identity to the 90mer substrates (10 matches out of 32 bases excluding the mismatch site and the base immediately 5′ to the mismatched G). Figure 9 shows that G·T mismatches located in CpG doublets in the 34mer oligonucleotides were good substrates for incision. No nicking of the matched duplex was observed, and no secondary cleavage product was seen in this experiment. This observation and the size of the radioactive reaction product, an 18mer, is consistent with cleavage immediately 5′ to the mismatched T since the substrate was labeled at the 5′ end of the T-containing strand. Of the five G·T mismatch-containing 34mer substrates, only those in which the mismatched G was 3′ to a C or a 5-meC were incised efficiently. Cleavage at G·T mismatches in the 34mer duplexes exhibited a hierarchy identical to that seen with the 90mer oligonucleotides. No nicking was detected when the mismatched G was immediately preceded on its 5′ side by G. Densitometric analysis indicated that the substrate in which the G was 3′ to a T or an A was cleaved to about 10% of maximum. Thus, efficient incision is confined to those G·T mismatches that are

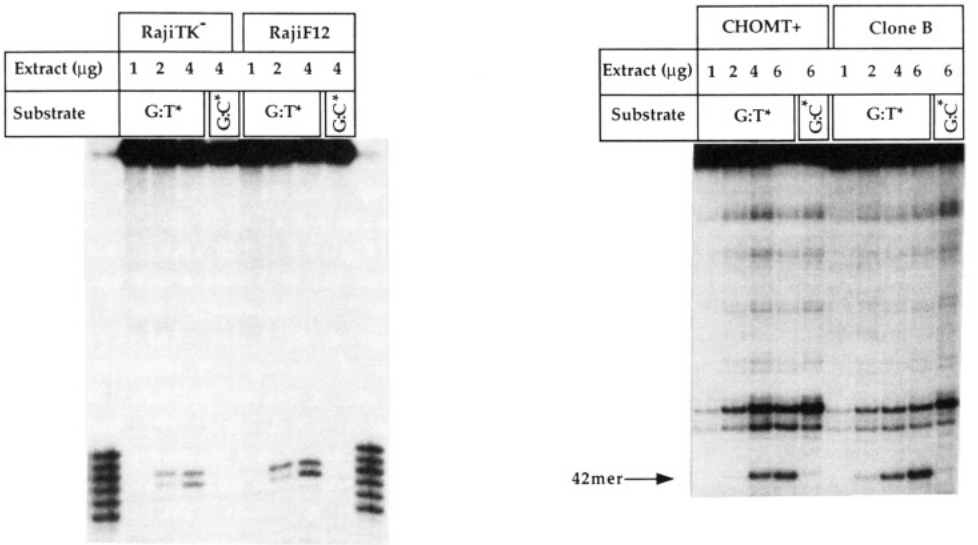

FIGURE 7: G·T mismatch cleavage by extracts of normal and G·T binding-deficient cells. (a) RajiTK⁻ and RajiF12. Matched or G·T mismatched 90mer duplexes (Figure 1 (parts a and b)) were 3′-end-labeled (*) in the pyrimidine-containing strand and incubated under standard reaction conditions with extracts of either RajiTK⁻ and RajiF12 as indicated. Incubation was for 16 h at 30 °C, and the products were analyzed as described in Materials and Methods. (b) CHOMT⁺ and Clone B. Extracts of CHOMT⁺ or Clone B cells were assayed as described above with the modifications described in Materials and Methods. The position of migration of a 42mer is shown arrowed.

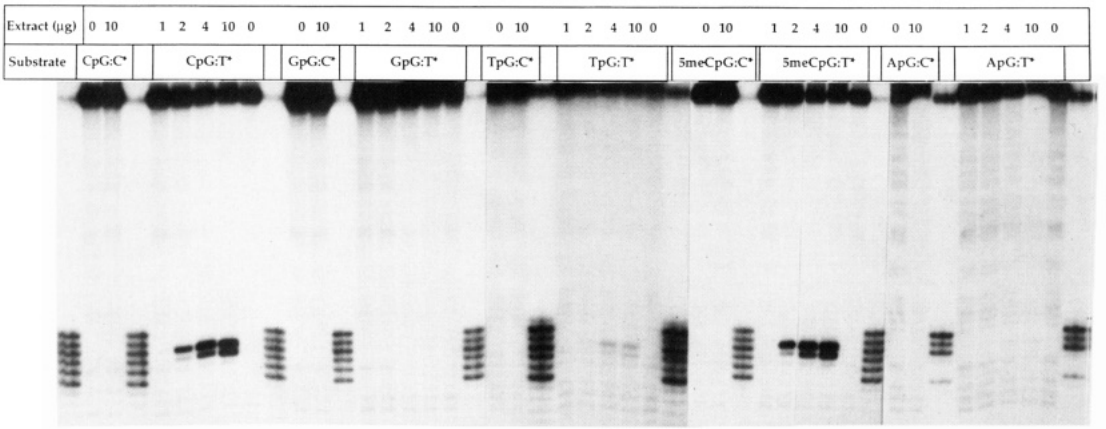

FIGURE 8: Sequence specificity of G:T mismatch cleavage in duplex 90mers. Duplex 90mer oligonucleotides contained either a G:C base pair or a single G·T mismatch at position 42. The matched or mismatched base pairs were in a CpG, GpG, TpG, 5-meCpG, or an ApG sequence at positions 41 and 42 (Figure 1b–f) as indicated. These duplexes were 3′ end-labeled in the strand indicated (*) and incubated with the amounts of HeLaMR extract shown for 16 h at 30 °C. Products were analyzed as described in Materials and Methods.

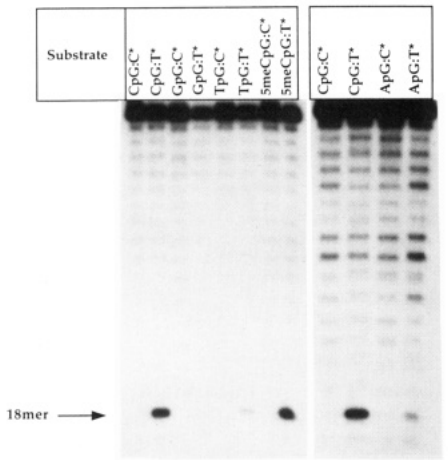

FIGURE 9: Sequence specificity of G:T mismatch cleavage in duplex 34mers. Duplex 34mer oligonucleotides containing a single G·T mismatch in a CpG, GpG, TpG, 5-meCpG or an ApG sequence at positions 15 and 16 on the top strand (Figure 2b–f) or the perfectly matched controls as indicated were incubated with the 10 μg HeLaMR extract for 8 h at 30 °C. Substrates were 5′-end-labeled in the lower strands as written in Figure 2. Products were analyzed as described in Materials and Methods. The position of migration of a 18mer is shown arrowed.

located in methylated or unmethylated CpG sequences in DNA. Closely similar extents of preferential incision were observed in two unrelated substrates indicating that incision is independent of the overall sequence context in which the CpG sequence is located.

DISCUSSION

The suggestion (Jiricny et al., 1988) that the mammalian G·T mismatch binding reaction might direct the thymine-DNA glycosylase to its site of action was attractive because it implied a mechanism for recognizing inappropriate thymine bases. However, unpublished data (cited in Hughes and Jiricny, 1992) suggest that the two pathways might be separate. The mutator phenotype of the alkylation resistant variant cell lines is associated with a deficiency in G·T-mismatch binding (Branch et al., 1993) and presents an ideal means of testing this possibility. Our finding that extracts of the variant cells perform normal levels of incision at G·T mismatches indicates that this binding reaction is indeed probably not required for incision. Further confirmation of this point comes from the observation that formation of the G·T-specific protein: oligonucleotide complex does not exhibit a sequence dependence. In particular, the presence of T, G, C, 5-meC, or A immediately 5′ to the mismatched G does not affect the G·T binding reaction. This relaxed specificity is in contrast to the G·T mismatch incision activity which exhibits a substantial preference for mismatches located in CpG sequences.

Three mismatch incision activities have been reported in human cell extracts (Yeh et al., 1991). One of these is specific for A·G mispairs, does not cleave G·T mismatches and thus resembles the *E. coli* MutY protein. The high efficiency of cleavage that we observe at $\frac{\text{CpG}}{\text{GpT}}$ and $\frac{\text{5meCpG}}{\text{GpT}}$ sequences effectively rules out the involvement of the A·G nicking activity in this process. A second activity, designated "all type" nicks all single base mispairs and appears to be considerably influenced by neighboring DNA sequences. The cleavage that we observed, unlike the "all type" incision, did not occur to any measurable extent on either strand of an A·C mispair and was instead preferentially directed to G·T mismatches in CpG sequences. This preference was observed in two substrates of unrelated overall sequence and this effectively rules out a substantial role for the "all type" activity.

The G·T mismatch incision activity we detect in replication extracts of HeLaMR, RajiTK$^-$, CHOMT$^+$, and the resistant variant cells is most probably the same as those previously reported in whole-cell or nuclear extracts of human cell lines (Wiebauer & Jiricny, 1989; Yeh et al., 1991; Sibghat-Ullah & Day, 1993). The cleavage reaction displays characteristic slow kinetics, the extent of incision increasing for up to 10 h, and is confined to the T-containing strand of G·T mismatches. One minor difference is that our data support the assertion that initial nicking occurs 5′ to the mismatched T (Sibghat-Ullah & Day, 1993) and not 3′ as originally suggested (Wiebauer & Jiricny, 1989). The kinetics of appearance of the reaction products is consistent with the proposed initiation of repair by a relatively slowly acting thymine-specific DNA glycosylase and subsequent cleavage of the DNA by an AP endonuclease. Incision 5′ to an AP site is consistent with the specificity of the major mammalian AP endonucleases (Doetsch & Cunningham, 1990).

The most probable source of G·T mismatches is *via* deamination of 5meC which is found almost exclusively in CpG dinucleotides in mammalian DNA (Bird, 1986). Our data indicate that cleavage of mismatched thymine residues is most probably confined to $\frac{\text{CpG}}{\text{GpT}}$ and $\frac{\text{5meCpG}}{\text{GpT}}$ mismatches. G·T mispairs in other sequence contexts are poorly incised. A mismatched T within CpG sequences can be regarded as deriving from a deamination event in a methylated DNA

strand. Initiation of the short patch mismatch correction pathway will therefore be confined to G·T mismatches that arise by deamination in methylated DNA. In this regard, short patch correction of G·T mismatches by mammalian cells resembles the vsp pathway of *E. coli* which is initiated by an endonuclease that specifically recognizes the sites of cytosine methylation by the *dcm* methylase (Hennecke et al., 1991). The inability of human cell extracts to nick efficiently at T residues opposite G positioned 3′ to G, to A, or to T suggests that this particular repair pathway is confined to correction of deaminated 5-meC and may not participate in the correction of G·T mismatches formed during DNA replication or recombination. Our conclusions are based on experiments *in vitro* and are in agreement with the expected preference for initiation of correction of deaminated 5meC. It remains unclear why repair of heteroduplex DNA transfected into mammalian cells exhibits a strong bias in correction of G·T to G:C that is apparently independent of the sequence context of the G·T mispair (Brown & Jiricny, 1987; Heywood & Burke, 1990).

A recent report (Sibghat-Ullah & Day, 1993) that whole-cell extracts also fail to incise an oligonucleotide in which mismatched G is located 3′ to a G is in agreement with our findings. When the mispair was placed in an unusually G+C rich region of the same oligonucleotide that contained CpG dinucleotides closely adjacent to and flanking the mismatch, the authors concluded that incision was independent of the base 5′ to the mismatched G. The reported competition experiments do not exclude the possibility of preferential incision at CpG located mismatches, and, indeed, much of the data appear to support this possibility. Using two unrelated substrates which do not contain unusual sequences we have obtained unequivocal evidence for preferential incision at G·T mismatches located at potential sites of 5-meC deamination. The data were obtained by direct titration of incision activity on each of the heteroduplex substrates and do not rely on interpreting indirect experiments measuring competition by unlabeled oligonucleotides. Our results clearly indicate that the incision of G·T mispairs measured in this assay is preferentially targeted to methylated CpG sequences. Since the properties of our incision reaction closely resemble the mismatched thymine-DNA glycosylase mediated initiation of short-patch repair of G·T mismatches in mammalian cell extracts (Wiebauer & Jiricny, 1989), we propose that repair of G·T mismatches by a short-patch pathway will occur preferentially in those sequences in which thymine bases arise *via* hydrolytic deamination.

It is probable that the short patch correction of G·T mispairs occurs predominantly, if not exclusively, in resting, fully methylated DNA. In the absence of specific "tagging" by a binding protein, the selective removal of thymine from mispairs presents a formidable problem of identification. Cleavage by cell extracts at mismatched T is extremely slow. Since AP endonucleases are abundant in cell extracts, the rate of cleavage is clearly limited by recognition and removal of the T. In contrast, the specific DNA glycosylases for deaminated cytosine (uracil) or adenine (hypoxanthine) act rapidly to remove these non-DNA bases (Dianov & Lindahl, 1991; Wittwer et al., 1989). The slow rate of cleavage at T is seen with extracts of mammalian cells prepared by several different methods and therefore probably accurately reflects a sluggish reaction rate *in vivo* rather than the artifactual loss of a limiting or labile component during extraction. The slow repair of mispaired T may at least partly explain the observation that 5-meC is a mutational hot-spot in DNA and the phenomenon of CpG suppression in which cytosine methylation sites have been lost, with a compensatory increase in TpG and CpA, during evolution (Russell et al., 1976). These phenomena suggest that the repair of deaminated 5-meC *in vivo* is also relatively inefficient.

In summary, there is a pathway that corrects G·T mispairs that arise by deamination of 5-meC in DNA in mammalian cells. Specificity of the pathway is conferred by the considerable preference of the G·T mismatch cleavage reaction for G·T mismatches in CpG or 5meCpG sequences. The dependence of the incision reaction on CpG dinucleotides is analogous to the sequence requirements of the Vsr endonuclease of *E. coli*, and this targeting may at least partly overcome the problem of recognition of ectopic thymine residues in DNA. Mismatches arising during replication or recombination should be repaired by a pathway that is independent of the local sequence context of the mispair. This criterion is fulfilled by the broad specificity of the G·T binding activity which recognizes the mispair equally in every context we examined. This indicates that this G·T binding reaction might be involved in a long-patch mismatch repair pathway (Holmes et al., 1990; Varlet et al., 1990; Thomas et al., 1991) that acts during recombination or replication, although its participation in removal of deaminated 5-meC in hemimethylated DNA by long-patch repair is not ruled out. Participation of G·T mismatch binding in a pathway of mismatch correction is consistent with the observed mutator phenotype of the variant Raji, CHO, and MT1 cell lines.

ACKNOWLEDGMENT

We thank Dr. G. Aquilina for cell lines, P. Branch for her help with the binding experiments, and the ICRF Oligonucleotide Synthesis Group for oligonucleotides.

REFERENCES

Aquilina, G., Frosina, G., Zijno, A., DiMuccio, A., Dogliotti, E., Abbondandolo, A., & Bignami, M. (1988) *Carcinogenesis 9*, 1217–1222.
Bird, A. (1986) *Nature 321*, 209–213.
Branch, P., Aquilina, G., Bignami, M., & Karran, P. (1993) *Nature 362*, 652–654.
Brown, T. C., & Jiricny, J. (1987) *Cell 50*, 945–950.
Brown, T. C., & Jiricny, J. (1988) *Cell 54*, 705–711.
Dianov, G., & Lindahl, T. (1991) *Nucleic Acids Res. 19*, 3829–3833.
Doetsch, P. W., & Cunningham, R. P. (1990) *Mutat. Res. 236*, 173–201.
Hennecke, F., Kolmar, H., Bründl, K., & Fritz, H.-J. (1991) *Nature 353*, 776–778.
Heywood, L. A., & Burke, J. F. (1990) *Mutat. Res. 236*, 59–66.
Holliday, R. (1964) *Genet. Res. 5*, 282–304.
Holmes, J., Clark, S., & Modrich, P. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 5837–5841.
Hughes, M. J., & Jiricny, J. (1992) *J. Biol. Chem. 267*, 23876–23882.
Jiricny, J., Hughes, M., Corman, N., & Rudkin, B. B. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 8860–8864.
Jones, M., Wagner, R., & Radman, M. (1987) *J. Mol. Biol. 194*, 155–159.
Kat, A., Thilly, W. G., Fang, W. H., Longley, M. J., Li, G. M., & Modrich, P. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 6424–6428.
Li, J. J., & Kelly, T. J. (1985) *Mol. Cell. Biol. 5*, 1238–1246.
Lieb, M. (1987) *J. Bacteriol. 169*, 5241–5246.
Lieb, M. (1991) *Genetics 128*, 23–27.
Michaels, M. L., Cruz, C., Grollman, A. P., & Miller, J. H. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 7022–7025.

Modrich, P. (1989) *J. Biol. Chem. 264*, 6597–6600.
Radman, M., & Wagner, R. (1986) *Annu. Rev. Genet. 20*, 523–538.
Russell, G. J., Walker, P. M. B., Elton, R. A., & Subak-Sharp, J. H. (1976) *J. Mol. Biol. 108*, 1–23.
Sibghat-Ullah, & Day, R. S. (1992) *Biochemistry 31*, 7998–8008.
Sibghat-Ullah, & Day, R. S. (1993) *Nucleic Acids Res. 21*, 1281–1287.
Stephenson, C., & Karran, P. (1989) *J. Biol. Chem. 264*, 21177–21182.
Thomas, D. C., Roberts, J. D., & Kunkel, T. A. (1991) *J. Biol. Chem. 266*, 3744–3751.
Varlet, I., Radman, M., & Brooks, P. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 7883–7887.
Wiebauer, K., & Jiricny, J. (1989) *Nature 339*, 234–236.
Wiebauer, K., & Jiricny, J. (1990) *Proc. Natl. Acad. Sci. U.S.A. 78*, 5842–5845.
Wittwer, C. U., Bauw, G., & Krokan, H. E. (1989) *Biochemistry 28*, 780–784.
Yeh, Y. C., Chang, D. Y., Masin, J., & Lien-Lu, A. (1991) *J. Biol. Chem. 266*, 6480–6484.