Interaction of MutS Protein with the Major and Minor Grooves of a Heteroduplex DNA*

(Received for publication, January 10, 1997, and in revised form, March 12, 1997)

Indranil Biswas and Peggy Hsieh‡

From the Genetics and Biochemistry Branch, NIDDK, National Institutes of Health, Bethesda, Maryland 20892-1810

*Thermus aquaticus* MutS protein is a DNA mismatch repair protein that recognizes and binds to heteroduplex DNAs containing mispaired or unpaired bases. Using enzymatic and chemical probe methods, we have examined the binding of *Taq* MutS protein to a heteroduplex DNA having a single unpaired thymidine residue. DNase I footprinting identifies a symmetrical region of protection 24–28 nucleotides long centered on the unpaired base. Methylation protection and interference studies establish that *Taq* MutS protein makes contacts with the major groove of the heteroduplex in the immediate vicinity of the unpaired base. Hydroxyl radical and 1,10-phenanthroline-copper footprinting experiments indicate that MutS also interacts with the minor groove near the unpaired base. Together with the identification of key phosphate groups detected by ethylation interference, these data reveal critical contact points residing in the major and minor grooves of the heteroduplex DNA.

DNA mismatch repair is crucial for maintaining the integrity of the genome. This DNA repair pathway corrects mispaired or unpaired bases that arise by replication errors, by physical damage to bases (e.g. deamination of 5-methylcytosine), and by the formation of heteroduplex DNA during homologous recombination between similar but not identical sequences (reviewed in Ref. 1). In addition, mismatch repair serves to modulate the formation of heteroduplex DNA during homologous recombination between evolutionarily divergent sequences (2, 3). More recently, mismatch repair has been implicated in transcription-coupled nucleotide excision repair and in G2 cell-cycle checkpoint control (reviewed in Ref. 4). Numerous genetic and biochemical studies have established the existence of mismatch repair in eukaryotes that bears overall similarity to that seen in prokaryotes including the identification of several eukaryotic MutS and MutL homologs. The recent demonstration that mutations in human MutS and MutL homologs are the underlying defect in many hereditary diseases indicates this fact.

The most thoroughly studied mismatch repair system is methyl-directed mismatch repair in *Escherichia coli* (reviewed in Refs. 1 and 6). *E. coli* MutS protein recognizes and binds to mismatches and small insertion/deletion mutations of 1–4 nucleotides. In the presence of ATP, MutL protein forms a complex with MutS bound to a heteroduplex DNA, and together they activate MutH, an endonuclease that incises the DNA at hemimethylated d(GATC) sites, thereby directing excision and repair synthesis to the newly synthesized (and transiently undermethylated) DNA strand. The nature of the strand discrimination signal in other prokaryotes and in eukaryotes where no equivalent dam methylation systems have been found is unclear, but it may involve single-strand breaks or ends (reviewed in Ref. 1).

The means by which MutS proteins discriminate between damaged and undamaged DNA poses an intriguing problem in protein-DNA recognition. The mismatch repair machinery of *E. coli* discriminates among different mismatches (reviewed in Ref. 6). G:T, A:C, A:A, and G:G mismatches are efficiently repaired, whereas the repair efficiencies of T:T, T:C, and A:G mismatches are dependent on sequence context. In contrast, C:C mismatches are very poor substrates for the mismatch repair pathway. X-ray diffraction and NMR studies indicate that, to a first approximation, most base pair mismatches are relatively well accommodated within the DNA helix and introduce only minor distortions (reviewed in Refs. 7 and 8). Nevertheless, these localized conformational changes can, in some cases, be detected biochemically, e.g. by alterations in the binding mode of the minor groove reporter neocarzinostatin (9) or by hypersensitivity to osmium tetroxide in the case of a G:T mismatch (10). Several studies have concluded that efficient repair involves a DNA bulge within the helix. Thus, intrahelical wobble base pairs (e.g. G:T, A:C, A:A, and G:G) may be repaired more readily than looped out base pairs (e.g. C:C, C:T, and T:T) (11–13). The recognition of DNA bulges is complicated by the fact that unpaired bases are, for the most part, stacked within the helix, although in some cases, unpaired pyrimidines have been observed to be extrahelical, and heteroduplexes containing unpaired bases are sometimes associated with compensatory kinks or bends (reviewed in Ref. 14).

To begin to address the molecular mechanism of substrate recognition by MutS protein, we have initiated biochemical studies of heteroduplex DNA binding by a thermostable MutS protein from *Thermus aquaticus*. This protein binds specifically to heteroduplex DNAs containing mispaired or unpaired bases over a broad temperature range up to 70 °C (15). Here, we utilize enzymatic and chemical probes to identify regions of the heteroduplex DNA that are in close association with *Taq* MutS protein. Our data reveal that critical contact points reside in the immediate vicinity of an unpaired base in both the major and minor grooves of a heteroduplex DNA.

**MATERIALS AND METHODS**

Purification of the *T. aquaticus* MutS Protein—*E. coli* strain BL21(DE3) (Novagen) transformed with pETMutS was grown at 37 °C in an air shaker or fermenter and induced for 4 h with 1 mM isopropyl-1-thio-β-D-galactopyranoside as described previously (15). The cells
MutS-Heteroduplex Interactions

were resuspended in lysis buffer, 20 mM KPO4 (pH 8.0), 100 mM NaCl, 0.1 mM dithiothreitol, 0.1 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, and 10 μg/ml each peptatin and leupeptin (30 ml of lysis buffer/liter of culture) and stored at 20 °C. Cells were thawed and disrupted by sonication followed by centrifugation for 60 min at 100,000 × g in a Beckman Ti-60 rotor. The supernatant was heated at 70 °C for 40 min and centrifuged at 16,000 × g for 20 min. The supernatant was brought to 65% saturated (NH4)2SO4 and centrifuged for 20 min at 16,000 × g. The pellet was resuspended in phosphate buffer (20 mM potassium phosphate (pH 8.0), 100 mM NaCl, 10% (v/v) glycerol, 0.1 mM dithiothreitol, 0.1 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, and 100 μg/ml each peptatin and leupeptin plus the recommended concentration of a protease inhibitor mixture (Boehringer Mannheim) and extensively dialyzed against the same buffer. The protein sample was then applied to a Mono Q HR 10/10 column (Pharmacia Biotech Inc.) as described (15). T. aquaticus MutS eluted at approximately 260 mM NaCl. Fractions containing MutS were pooled and dialyzed against a HEPES buffer containing 20 mM HEPES (pH 7.8), 50 mM NaCl, 10% (v/v) glycerol, 0.1 mM EDTA, 0.1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride, and 100 μg/ml each peptatin and leupeptin. The dialyzed sample was applied to a hydroxyapatite column (Bio-Rad), bed volume 10 ml, equilibrated in HEPES buffer. The column was developed with a 80-ml linear gradient of 0–120 mM potassium phosphate in HEPES buffer. MutS eluted at 60 mM potassium phosphate. MutS protein was then applied to a Hi-Prep Sephacryl S-300 column (Pharmacia Biotech Inc.), bed volume 26 ml, equilibrated in HEPES buffer. The protein was concentrated by passage over a second Mono Q HR 10/10 column as described above. One liter of culture yielded about 2 mg of protein in the final fraction.

Preparation of Duplex Substrates—Synthetic DNA substrates used in gel mobility shift assays were prepared as described previously (15). For footprinting assays, substrates were derived from restriction endonuclease fragments. Oligonucleotide duplexes corresponding to –12 to +11 position in the top strand and –9 to +15 in the bottom strand (see Fig. 2) were cloned into the SaI site of pBlueScript IIISK (Stratagene). Recombinant plasmids were digested with EcoRI and KpnI, and the corresponding fragments were purified on a 10% native polyacrylamide gel. Complementary strands were separated on 6% denaturing polyacrylamide gels. The resulting single-strand DNA was 5′-end-labeled with [γ-32P]ATP (6000 Ci/mmol; DuPont NEN) and T4 polynucleotide kinase (New England BioLabs) and annealed with an equimolar concentration of an unlabelled complementary strand to generate a homoduplex or heteroduplexes containing an unpaired thymidine residue or a G:T mismatch.

Gel Mobility Shift Assay—DNA binding assays were carried out in 20 μl of 20 mM HEPES, pH 7.8 (25 °C), 0.1 mM dithiothreitol, 0.1 mM EDTA, 5% glycerol, and 5 mM MgCl2. The reactions were initiated by the addition of 0.5 pmol of a 5′-32P-labeled 37-base pair duplex DNA (A:T, G:T, or Δ1) and various amounts of MutS protein as indicated. After incubation at 60 °C for 15 min, the reactions were electrophoresed on a 4.8% native polyacrylamide gel containing 1× TBE and 10 mM MgCl2. Electrophoresis was carried out at room temperature in 1× TBE containing 5 mM MgCl2. Quantitation was on a Fuji phosphor imager.

DNase I Footprinting—Approximately 1.8 μg of MutS (20 pmol) was incubated with 35 fmol of 5′-32P-labeled heteroduplex DNA fragments for 15 min at 60 °C as described above with the addition of 20 μg/ml poly(dI-dC) (Pharmacia). DNase I footprinting was carried out in the presence of 0.025 or 0.10 units of DNase I according to the manufacturer’s directions (Stratagene). The phenol-chloroform-extracted DNA was dissolved in 4 μl of 90% formamide/dye mixture and electrophoresed on an 18% denaturing polyacrylamide gel.

Methylation Protection and Interference—For methylation protection assays, DNA binding was carried out as described for DNase I footprinting assays except that 50 fmol of heteroduplex DNA and 10 pmol of MutS were used. After binding at 60 °C, 1 μl of salmon sperm DNA (10 mg/ml) and 2 μl of a 1:10 dilution of dimethyl sulfate (Aldrich) were added and incubated at room temperature for 2 min. Reactions were terminated by the addition of 50 μl of a stop solution (1.5 mM sodium acetate, 1 mM mercaptoethanol, and 250 μg/ml tRNA). The DNA was ethanol-precipitated, lyophilized, and incubated with 6.5 μl of 0.1 M Fe(EDTA)2 (Aldrich) and CuSO4 were added, resulting in a final concentration of 200 μM 5-phenyl-1,10-phenanthroline and 45 μM copper ion. After 30 min, 6.5 μl of 28 μM 2,9-dimethyl-1,10-phenanthroline (Aldrich) were added to quench the reaction. Samples were prepared for 12% sequencing gels as described above.

Hydroxyl Radical Footprinting—Heteroduplex DNA binding was carried out as described for methylation protection assays except that glycerol and poly(dI-dC) were omitted. The preparation of the (FeEDTA)2 complex, incubation with MutS-heteroduplex complexes, and preparation of DNA fragments for electrophoresis on 12% sequencing gels were described previously (17).

Ethylation Interference Assay—5′-End-labeled heteroduplex DNA (300 pmol) was ethylated in the presence of ethylnitrosourea (Sigma) as described previously (18). Binding reactions were performed as in methylation interference experiments except that 120 fmol of heteroduplex DNA and 10 pmol of MutS were added to each reaction. Protein-bound and free DNA fractions were separated on native polyacrylamide gels as described for methylation interference assays. Strand cleavage in NaOH was performed as described (18), and the samples were analyzed on 12% denaturing polyacrylamide gels as described above.

Osmium Tetroxide Modification—DNA binding was carried out as described above in the presence of poly(dI-dC). After binding at 60 °C for 10 min, freshly diluted OsO4 (Sigma) and bis-pyridyl in 50% ethanol (Sigma) were added to a final concentration of 1 mM and 2.5 mM, respectively, and incubated for 30 s at 60 °C. The reaction mixtures were desalted on G50 spin columns (Pharmacia). Strand cleavage and analysis on 12% denaturing polyacrylamide gels were as described for the methylation protection assays.

RESULTS

Size Exclusion Chromatography of Taq MutS Protein—T. aquaticus MutS protein was overproduced in E. coli and purified to apparent homogeneity as described under “Materials and Methods.” Molecular mass standards and their Stokes radii are as follows: blue dextran, thyroglobulin (669 kDa, 85.0 Å), apoferritin (443 kDa, 61.0 Å), β-amylose (200 kDa), aldolase (158 kDa, 48.1 Å), alcohol dehydrogenase (150 kDa), and bovine serum albumin (67 kDa, 35.5 Å). Right, 2.0 μg of MutS protein was resolved on a 10% SDS-polyacrylamide gel electrophoresis followed by Coomassie Blue staining. Positions of molecular mass standards (Bio-Rad) are indicated on the left.

FIG. 1. Size exclusion chromatography of Taq MutS protein. Left, Taq MutS protein was chromatographed on a Sephacryl S-300 column as described under “Materials and Methods.” Molecular mass standards and their Stokes radii are as follows: blue dextran, thyroglobulin (669 kDa, 85.0 Å), apoferritin (443 kDa, 61.0 Å), β-amylose (200 kDa), aldolase (158 kDa, 48.1 Å), alcohol dehydrogenase (150 kDa), and bovine serum albumin (67 kDa, 35.5 Å). Right, 2.0 μg of MutS protein was resolved on a 10% SDS-polyacrylamide gel electrophoresis followed by Coomassie Blue staining. Positions of molecular mass standards (Bio-Rad) are indicated on the left.
Sephacryl S-300 reflects a higher order structure for the free protein in solution or an asymmetrical shape for the protein monomer.

Relative Binding Affinities of MutS for Heteroduplex DNA—The binding of MutS protein to heteroduplex DNAs containing either a single base insertion of a thymidine residue (Δ1), a single G:T mismatch (G:T), or a perfect homoduplex (A:T) was examined in gel mobility shift assays. The DNA substrates used for these studies and for the chemical footprinting studies described below are shown in Fig. 2.

DNA binding assays utilizing a heteroduplex containing a single unpaired thymidine residue were carried out in the presence of 5 mM MgCl₂ at 60 °C for 15 min in the presence of 25 nM ³²P-labeled Δ1 heteroduplex DNA made by annealing synthetic deoxyribonucleotides 36–37 nucleotides long mapping to positions −18 to +18 and containing one unpaired thymidine Δ1, a G:T mismatch, or an A:T base pair at position 0.

To assess the relative affinities of Taq MutS protein for an insertion/deletion substrate versus a G:T mismatch substrate, we performed a competition assay in which binding of Taq MutS protein to 1 pmol of ³²P-labeled Δ1 heteroduplex was challenged by the presence of varying concentrations of cold Δ1 or G:T competitor DNA (Fig. 3B). Loss of binding to the ³²P-labeled substrate was monitored in a gel mobility shift assay. Whereas an equimolar concentration of cold competitor Δ1 DNA reduced binding to labeled DNA to almost one-third (37% bound) of the level in the absence of any competitor, a 200-fold excess of G:T competitor DNA was required to reduce binding to a similar extent (43.5% bound). Consistent with the lower affinity for a G:T mismatch compared with an insertion/deletion substrate, complexes containing a G:T heteroduplex bound to MutS protein were detected in gel mobility shift assays only when a large excess of Taq MutS protein (500 nM) was incubated with 25 nM of labeled G:T heteroduplex (data not shown).

Filter binding assays were also consistent with a higher affinity for the Δ1 substrate than for the G:T substrate (data not shown). As we have described previously, Taq MutS has no significant affinity for homoduplex (A:T) DNA under conditions where heteroduplexes are readily bound (15).

DNase I Footprinting of a MutS-Heteroduplex Complex—Taq MutS protein was incubated with a ³²P-labeled Δ1 heteroduplex containing a single unpaired thymidine residue. This DNA substrate was 76 nucleotides long on the top strand and 67 nucleotides long on the bottom strand (see Fig. 2). We have previously shown that heteroduplex DNA binding by Taq MutS protein exhibits an absolute requirement for Mg²⁺ (15); therefore, all DNA binding was carried out in the presence of 5 mM MgCl₂. Under the conditions used, more than 95% of the heteroduplex DNA was bound by MutS protein (see "Materials and Methods").
MutS-Heteroduplex Interactions

by cleavage with pyrrolidine. The resulting DNA fragments were electrophoresed on a denaturing polyacrylamide gel (Fig. 5A).

Protection of guanine residues in either the top or bottom strand of the heteroduplex from modification by dimethyl sulfate was restricted to those guanine residues immediately flanking the base insertion (Fig. 5C). Protection was observed at positions −1 and +1 on the top strand and −2 and +2 on the bottom strand (numbering is with respect to the unpaired thymidine). The extent of protection of the −2 guanine was variable and weaker than that observed for the +2 guanine on the same strand. Nearby guanines at −3 on the top strand and −6 on the bottom strand were unaffected. Enhanced modification of guanine residues was observed in both the top and bottom strands but was limited to the guanine residue nearest the protected region, position +3 on the top strand and +5 on the bottom strand. Both hypersensitive sites were positioned to one side of the lesion (to the right as drawn in Fig. 3). Methylation protection experiments shown in Fig. 5A were carried out with a protein monomer-DNA molar ratio of 200:1. At a molar ratio of 10:1, the identical pattern of methylation protection was obtained (data not shown).

Interference of MutS binding by methylation of the N-7 position of guanine residues in the Δ1 heteroduplex was also examined. The 32P-labeled Δ1 heteroduplex was subjected to limited modification by dimethyl sulfate as described under “Materials and Methods.” The modified Δ1 heteroduplex was then incubated with Taq MutS protein under conditions in which approximately 50% of the DNA was bound. Free and bound DNAs were separated on nondenaturing polyacrylamide gels and excised. Following strand scission in pyrrolidine, the free and bound DNA fragments were analyzed on a denaturing polyacrylamide gel (Fig. 5B). Consistent with the methylation protection studies described above, only methylation of those guanine residues in the immediate vicinity of the lesion interfered with MutS binding. Thus guanine residues +1 and −1 on the top strand and −2 and +2 on the bottom strand are over-represented in the “free” DNA population and underrepresented or absent in the “bound” population. Two weak sites of methylation interference reside at +8 on the top strand and +7 on the bottom strand; their significance is unknown.

The results of the methylation protection and interference studies clearly demonstrate binding of the Taq MutS protein to the major groove of a heteroduplex containing a single base insertion mutation. This interaction is confined to those residues in the immediate vicinity of the lesion.

1,10-Phenanthroline-Copper Footprinting of a MutS-DNA Complex—Possible interactions between MutS and the minor groove of the Δ1 heteroduplex were probed with 1,10-phenanthroline-copper ion (16). The 1,10-phenanthroline-copper-ion forms a coordination complex with the minor groove of DNA, with the initial site of oxidative attack being the C-1′ hydrogen of the deoxyribose. Any protein that blocks access to the C-1′ hydrogen either directly or indirectly or that induces a structural distortion of the minor groove that precludes access to this position will prevent cleavage by this reagent.

Taq MutS protein was incubated with the 32P-labeled Δ1 heteroduplex as described above, followed by modification and strand scission in the presence of phenanthroline-copper. As shown in Fig. 6, chemical probing of the Δ1 substrate in the absence of MutS protein (∼MutS) reveals the presence of hyperreactive sites in both strands that map to the immediate vicinity of the unpaired base. These hyperreactive sites indicate distortion of the heteroduplex in the vicinity of the bulge. In contrast, in the presence of Taq MutS protein, a pattern of symmetrical protection from cleavage by phenanthroline-cop-
per is readily evident. Protection extends from residue −2 to +3 on the top strand and from −3 to +3 on the bottom strand. This region corresponds to the minor groove containing the unpaired thymidine. As expected, the cleavage pattern by phenanthroline-copper is unchanged in the presence or absence of MutS protein in the case of a homoduplex (A:T) control. These results indicate that upon binding of the heteroduplex by Taq MutS protein, the minor groove containing the unpaired thymidine is shielded from modification. Such a footprint may result from direct contacts by the MutS protein in the minor groove or from distortions of the minor groove induced by MutS protein.

Hydroxyl Radical Footprinting—To probe possible interactions involving Taq MutS protein and the sugar-phosphate backbone of the Δ1 heteroduplex, we carried out hydroxyl radical footprinting of MutS-Δ1 complexes. The hydroxyl radical reacts in a non-sequence-specific fashion with the deoxyribose moiety, resulting in hydrogen atom abstraction at all five carbon atoms (21). Our data cannot distinguish among the different sites of attack; however, as discussed below, the pattern of hydroxyl radical-induced cleavage of a duplex DNA can yield information on the nature of the interaction. Taq MutS protein was incubated with the 32P-labeled Δ1 heteroduplex as described above. After complex formation, the reaction mixture was subjected to cleavage by hydroxyl radicals generated via the Fenton reaction by reduction of iron(II) with hydrogen peroxide (22).

Hydroxyl radical footprinting of MutS-DNA complexes revealed complete protection of three nucleotides in the vicinity of the unpaired base in both the top and bottom strands (Fig. 7A). The two regions of strong protection are offset in the 3′ direction by 3 nucleotides. In addition, adjacent regions of weaker protection were observed in both strands oriented to one side of the unpaired base (i.e. to the 3′ side of the lesion in the top strand). These results are consistent with equivalent or nearly equivalent interactions of MutS protein across the minor groove in the immediate vicinity of the lesion, the 3′ offset of the protected regions reflecting the right-handed helical nature of the heteroduplex (23). However, as discussed above, protein-induced distortions of the minor groove in this vicinity cannot be ruled out. In addition, the presence of the strong protection of the unpaired thymidine and flanking guanine residues and the weaker protection of a region on the bottom strand displaced in the 5′ direction are consistent with interactions involving the major groove of the heteroduplex as indicated by the methylation protection and interference results described above.

Ethylation Interference—Interaction between Taq MutS protein and phosphates in the Δ1 heteroduplex were investigated by ethylation interference experiments using ethylnitrosourea (18). The Δ1 heteroduplex was modified with ethylnitrosourea prior to incubation with Taq MutS protein under conditions in which 50% binding was achieved. Bound and free DNA fragments were separated on a native polyacrylamide gel, hydrolyzed in alkali, and fractionated on a denaturing polyacrylamide gel (Fig. 7B). Ethylation interference was observed at
sequence context as the heteroduplex containing a G:T mismatch in the same sequence context. Hypersensitive sites in the G:T substrate at guanine residues +3 and +5 on the top and bottom strands, respectively, are identical to those in the Δ1 substrate. The weaker protection for the G:T heteroduplex is consistent with the lowered affinity of Taq MutS protein for the G:T mismatch substrate compared with the Δ1 insertion/deletion substrate.

The complex formed between Taq MutS protein and the G:T heteroduplex was also probed with 1,10-phenanthroline-copper (data not shown). The results are summarized in Fig. 8B. Protection from phenanthroline-copper cleavage was observed on the top strand at positions 0 through +4. Protection was also observed on the bottom strand at positions −3 through +3. The phenanthroline-copper footprint of the G:T heteroduplex was very similar to that of the Δ1 heteroduplex (see Fig. 6). In both cases, protection from cleavage was observed in the regions flanking the lesion on both strands. In the case of G:T binding, hyperreactive residues were observed on both the top strand (positions −4, −2, and −1) and bottom strand (positions −4 and +4). By analogy to the case of E. coli RNA polymerase bound to the lac promoter (16) or Flp recombinase bound to a target DNA (17), these 1,10-phenanthroline-copper-hyperreactive sites may reflect localized unwinding of the DNA by Taq MutS protein, although this has not been directly shown. Alternatively, MutS-induced distortion of the heteroduplex may render the C1′ hydrogens more accessible to modification.

Probing the Complex with Osmium Tetroxide—The presence of single-stranded regions in complexes was probed with osmium tetroxide. Osmium tetroxide modifies thymidylate residues in single- or double-stranded DNA in which there is increased accessibility of the 5,6-double bond, e.g. when base stacking is absent or reduced (24, 25). As shown in Fig. 9, no detectable cleavage was observed in the case of an A:T homoduplex or a G:T heteroduplex in the presence or absence of MutS. In the case of the Δ1 substrate, however, a prominent OsO4 cleavage site corresponding to the unpaired thymidylate residue was observed in the absence of MutS protein. Upon binding by MutS, however, this hyperreactivity was completely abolished, and no new cleavage sites were observed. OsO4 modification at 37 °C instead of 60 °C of the G:T and Δ1 substrates in the presence or absence of MutS protein yielded identical results (data not shown). Thus, no region of single-stranded DNA was detected by this assay in the MutS-heteroduplex complex. Bhattacharyya and Lilley (10) previously reported that the thymidine of a G:T mismatch was reactive to OsO4. The difference in findings may be related to differences in sequence context. The possibility that G:T mismatches in different sequence contexts may adopt alternative conformations as judged by reactivity to OsO4 may explain why Taq MutS protein binds more weakly to the particular G:T substrate used in this study compared with the Δ1 substrate.

Methylation Protection of a Different Sequence Context—While mismatch repair is not targeted toward any specific sequence, the identity of flanking sequences affects mismatch repair in vivo (11, 26) and in vitro (27). Also, sequence context can affect the conformation of a base pair mismatch or bulge (reviewed in Refs. 14 and 28), and its effects are probably not limited to nearest neighbor considerations (see, for example, Refs. 13 and 29–31). We carried out methylation protection experiments with a Δ1 heteroduplex containing a single unpaired thymidine residue whose sequence differed in the vicinity of the unpaired base from that shown in Fig. 2 at positions

\[
\begin{align*}
5'.. & \text{ACGTCGACGCTAGGTTGCGTCTGTCGACCC..} 3' \\
3'.. & \text{TGGCAGCTCGACTGCCGCGCCGCGACGCGCTGG..} 5'
\end{align*}
\]

Fig. 6. 1,10-Phenanthroline-copper footprinting. 1,10-Phenanthroline-copper footprinting of a MutS-Δ1 complex was performed as described under "Materials and Methods." C, untreated heteroduplex; G, Maxim-Gilbert guanine sequencing of the top or bottom strand; Δ1, heteroduplex. The arrows indicate the position of unpaired thymidine residue. Lower panel, summary of the 1,10-phenanthroline-copper footprinting analysis.

several positions in both the top and bottom strands (see also Fig. 10). Not surprisingly, ethylation of phosphates located 3′ to all four critical guanine residues identified by methylation studies described above abolished binding (3′ cleavage of positions +1 and −1 on the top strand and +2 through +4 on the bottom strand). Ethylation of phosphates 3′ to positions +1 and −1 on the top strand and positions −2 through −5 on the bottom strand correspond to residues protected from hydroxyl radical attack and are consistent with interactions involving the opposing minor groove. In addition, modification of phosphates 3′ to residues +6 and +7 (top strand) and +6 through +8 (bottom strand) prevented binding of MutS protein.

Comparison of the Δ1 and G:T Heteroduplexes—Methylation and 1,10-phenanthroline-copper footprinting experiments with a heteroduplex containing a G:T mismatch in the same sequence context as the Δ1 heteroduplex described above (see Fig. 2) revealed that the interactions between MutS protein and the two heteroduplexes are very similar despite a greater affinity of the protein for the unpaired base compared with the G:T mismatch. As shown in Fig. 8A, complexes formed with the G:T heteroduplex yielded similar footprints with regard to methylation protection, these being protection of the two guanine residues flanking the mismatch on the top strand (positions −1 and +1) and protection of a flanking guanine on the bottom strand (position −2) as well as protection of the guanine involved in the G:T mismatch (position 0). The failure to observe protection of the −2 guanine position in the case of the G:T mismatch is consistent with the weak and somewhat variable protection seen in the case of the Δ1 substrate. Hypersensitive sites in the G:T substrate at guanine residues +3 and +5 on the top and bottom strands, respectively, are identical to those in the Δ1 substrate. The weaker protection for the G:T heteroduplex is consistent with the lowered affinity of Taq MutS protein for the G:T mismatch substrate compared with the Δ1 insertion/deletion substrate.
−2 through −4, +1, and +3 on the top strand. The sequence corresponding to the top strand in the region of the unpaired thymidine was 5’...GCTGATGTA...3’; the boldface thymidine is unpaired, and changed residues are underlined. The methylation protection data indicated no significant differences in the identity of protected or hyperreactive sites (data not shown). Flanking guanine residues were protected (position −21, top strand; position −12, bottom strand), while the guanine at position −15 on the bottom strand was hyperreactive. While this experiment only begins to address the issue of sequence context, it does indicate that at least for major groove contacts detected by dimethyl sulfate modification, there is no significant difference in the mode of binding by Taq MutS protein for two sequences that differ in their GC content and in the identity of one of the two base pairs that flank the identical insertion mutation.

**DISCUSSION**

Enzymatic and chemical footprinting identify regions of a heteroduplex DNA containing a single unpaired thymidine residue or a G:T mismatch that interact with a thermostable MutS mismatch repair protein. The DNase I footprint of the complex extends approximately three turns of the DNA helix more or less centered on the lesion. Chemical probes delineate protein-DNA interactions involving no more than one turn of the DNA helix in either direction flanking the lesion. Interactions involving a very limited number of nucleotides are consistent with the absence of a defined consensus sequence for mismatch recognition by MutS protein. Taq MutS protein contacts the major groove and the sugar-phosphate backbone of the heteroduplex in the region of an unpaired base. In addition, hydroxyl radical and phenanthroline-copper footprinting strongly suggest that MutS protein is also closely associated with the minor groove in the immediate vicinity of the unpaired base, although distortion of the minor groove as a consequence of binding elsewhere cannot be ruled out. Based on these data, we propose that Taq MutS makes extensive contacts with the periphery of the helix in the vicinity of the lesion.

Protein contacts involving the major groove of a Δ1 insertion/deletion heteroduplex in two different sequence contexts as well as a G:T heteroduplex are identified as a result of protection from methylation of neighboring guanine residues on either side of the lesion. Methylation interference implicates the same critical guanine residues in heteroduplex binding. Two regions of hydroxyl radical protection offset in the 5’ direction are also consistent with major groove interactions in the immediate vicinity of the lesion. Independent evidence for major groove contacts involving Taq MutS protein stem from photocross-linking studies in which the cross-linking moiety is located in the major groove of a heteroduplex DNA.\(^1\)

Close association of the Taq MutS protein with the opposing minor groove in the vicinity of the lesion is supported by hydroxyl radical and phenanthroline-copper footprinting results. Hydroxyl radical footprinting of the Taq MutS-Δ1 heteroduplex complex reveals a pair of completely protected regions in complementary strands offset by 3 nucleotides in the 3’ direction.

\(^1\) V. Malkov, I. Biswas, R. D. Camerini-Otero, and P. Hsieh, submitted for publication.
This pattern of strong protection from hydroxyl radical cleavage is characteristic of minor groove binding as first demonstrated for λ repressor (22) and subsequently for other proteins including E. coli IHF protein (32), TFIIID (33), and Flp recombinase (17). Although we cannot rule out MutS-induced distortion of the minor groove in the absence of minor groove binding, the pattern of hydroxyl radical cleavage does rule out alignment of Taq MutS protein along only one side of the DNA, since no pattern of periodic protection coinciding with the helical repeat of DNA was observed.

Binding of Taq MutS protein to a ΔI or G:T heteroduplex also affords complete protection from 1,10-phenanthroline-copper modification of approximately 6 residues on both strands centered on the lesion (Figs. 6 and 8B). Such extensive protection in conjunction with hydroxyl radical footprinting, discussed above, and ethylation interference, see below, argue strongly for protein contacts involving the minor groove.

Ethylation interference studies identify several phosphates in both strands of the heteroduplex as being important for binding and are consistent with the protein contacting both the major and minor grooves in the vicinity of the unpaired base (see Fig. 10). Ethylation of phosphates located 3' to every critical guanine residue identified in methylation protection/interference studies abolished MutS binding. Likewise, ethylation of phosphates 3' to every protected residue identified in hydroxyl radical footprinting experiments, with the exception of the unpaired thymidine, also abolished binding.
MutS-Heteroduplex Interactions

Our results confirm and extend previous studies concerning interactions between MutS proteins and heteroduplex DNA. DNase I protection and methidiumpropyl-EDTA interactions between MutS proteins and heteroduplex DNA are reminiscent of the binding interactions of another DNA repair protein, E. coli MutY, that recognizes G:A mismatches. Based on alkylation interference and substrate specificity studies, those authors concluded that MutY binds to both the major and minor grooves in the vicinity of the mismatch as well as to several critical phosphate groups.

Central to the issue of recognition of mispaired and unpaired bases by MutS proteins is the structure of the DNA heteroduplex containing such a lesion. Unfortunately, our current knowledge of DNA structure as it relates to mismatches or bulges is far from complete. However, NMR, x-ray diffraction, electrofootprinting, and thermodynamic studies reveal little structural similarity between DNA bulges and mismatches.

Both x-ray crystallographic and NMR studies reveal that the G:T mismatch assumes a “wobble” type hydrogen bonding scheme utilizing the major tautomers of the bases (reviewed in Refs. 28, 36, 37). In this structure, the thymine is displaced toward the major groove. Distortions of the G:T heteroduplex are not as extensive as for a bulged DNA (see below), although the mismatch does result in localized structural distortions including alterations in base stacking in adjacent residues in some instances (8). In addition, a G:T mismatch has been shown to alter the attack site of the minor groove intercalator neocarzinostatin (9), and the mispaired thymidine in a G:T mismatch can be hypersensitive to modification by osmium tetroxide (10).

Unpaired bases can be extrahelical or stacked within the helix. Unpaired purines are largely stacked within the DNA helix at all temperatures and in several different sequence contexts (38–41). Pyrimidine bulges can be either extrahelical or intrahelical, depending on sequence and temperature. In general, pyrimidine bulges are looped out at lower temperatures, i.e. 0 °C, and stacked within the helix at higher temperatures, i.e. 35 °C (30, 42–45). In some although not all cases, the presence of a DNA bulge confers a bend in the DNA helix (reviewed in Ref. 14). The extent of bending can be influenced by the identity of the nucleotides in the bulge (unpaired purines cause greater bending than unpaired pyrimidines), the size of the bulge, and the flanking sequence (10, 46, 47). While the conformation of the unpaired thymidine in the DNA substrate shown in Fig. 2 is unknown, it may be stacked within the helix, as has been established by NMR for another bulged heteroduplex consisting of one unpaired thymidine flanked by guanine residues (44). If the thymidine is stacked within the helix, it is still subject to hypermodification by 1,10-phenanthroline-copper and OsO₄ prior to being bound by MutS, and the presence of the unpaired base introduces distortion of neighboring residues as detected by phenanthroline-copper footprinting of naked heteroduplexes.

A possible mechanism for the recognition of insertion/deletion and mismatch lesions by MutS proteins is one that has been invoked for nucleotide excision repair to explain damage-specific recognition of a broad spectrum of DNA lesions that share no obvious structural similarities (reviewed in Refs. 48 and 49). In this scheme, recognition is mediated through protein-induced deformation of the DNA rather than direct recognition of structurally diverse substrates. Presumably, lesions that are targets for excision repair facilitate the formation of an altered conformation, thereby shifting the equilibrium in favor of the formation of a stable protein-DNA complex. Based on the work presented here and the work of others, we speculate that the DNA bound and distorts the heteroduplex DNA and utilizes extensive contacts with the periphery of the helix to gauge the extent to which a region of DNA can accommodate protein-induced strain. Both Mg²⁺ and ATP cofactors have been shown to modulate heteroduplex binding by MutS proteins (reviewed in Refs. 1 and 4), and such modulation may be an integral component of the target search by MutS. In this regard, it is noteworthy that E. coli MutS protein has been shown to form α structures in the presence of ATP, suggesting that it can translocate along the DNA (6).

This study provides new details on the interactions between MutS protein and a heteroduplex DNA. Problems to be pursued include the identification of regions of MutS protein that contact the major and minor grooves of the heteroduplex, the determination of protein-induced DNA distortion, and an examination of how contacts between MutS protein and the heteroduplex DNA are modulated during repair.

Acknowledgments—We are grateful to Dan Camerini-Otero, Wei Yang, and Oleg Voloshin for helpful comments on the manuscript. We thank George Foy for oligonucleotide syntheses and Linda Robinson for assistance.

REFERENCES
1. Modrich, P., and Lahue, R. (1996) Annu. Rev. Biochem. 65, 101–133
2. Rayssiguier, C., Thaler, D. S., and Radman, M. (1989) Nature 342, 396–401
3. Selva, E. M., New, L., Crouse, G. F., and Lahue, R. S. (1995) Genetics 139, 1175–1188
4. Kolodner, R. (1996) Genes & Dev. 10, 1433–1442
5. Kunkel, T. A. (1995) Curr. Biol. 5, 1091–1094
6. Modrich, P. (1991) Annu. Rev. Genet. 25, 229–253
7. Kennard, O. (1985) J. Biomol. Struct. & Dyn. 3, 205–226
8. Patel, D. J., Shapiro, L., and Hare, D. (1987) in Nucleic Acids and Molecular Biology (Ekkestein, F., and Lilley, D. M. J., eds) Vol. 1, pp. 70–84, Springer-Verlag, Berlin
9. Kappen, L. S., and Goldberg, I. H. (1992) Biochemistry 31, 9081–9089
10. Bhattacharyya, A., and Lilley, D. M. J. (1988) Nucl. Acids Res. 17, 6821–6840
11. Fazakerley, G. V., Quignard, E., Woisard, A., Guschlbauer, W., van der Marel, G., Struyven, B. J. H., Jones, M., and Radman, M. (1996) EMBO J. 5, 3697–3703
12. Su, S. S., and Modrich, P. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 5057–5061
13. Werniges, H., Steger, G., Riesen, D., and Fritz, H.-J. (1996) Nucl. Acids Res. 24, 3773–3790
14. Turner, D. H. (1992) Curr. Opin. Struct. Biol. 2, 334–337
15. Biswas, I., and Hsieh, P. (1996) J. Biol. Chem. 271, 5040–5048
16. Kowabara, M. D., and Sigman, D. S. (1987) Biochemistry 26, 7234–7238
17. Kimball, A., Kimball, M. L., Jayaram, M., and Tullius, T. D. (1995) Nucleic Acids Res. 23, 3009–3017
18. Siebenlist, U., and Gilbert, W. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 122–126
19. Laurent, T. C., and Killander, J. (1964) J. Chromatogr. 14, 317–330
20. Kurnit, D. M., and Dervan, P. B. (1990) Science 248, 847–850
21. Palaecek, E. (1992) Methods Enzymol. 212, 159–155
22. Nielsen, P. E. (1990) J. Mol. Recog. 1, 1–25
23. Jones, M., Wagner, R., and Radman, M. (1987) Genetics 115, 605–610
24. Jiricny, J., Su, S. S., Wood, S. G., and Modrich, P. (1988) Nucleic Acids Res. 16, 7843–7853
25. Kennard, O., and Salisbury, S. A. (1993) J. Biol. Chem. 268, 10701–10704
26. Kokniewicz, E., Roogna, V., Jones, C. R., and Gorenstein, D. G. (1989) Biochemistry 28, 9858–8572
27. Kohn, F., Zaltzman, P. A., Hager, M., and Itakura, K. (1984) Biochemistry 23, 7469–7473
28. Kohn, F., Zaltzman, P. A., Hager, M., and Itakura, K. (1984) Biochemistry 23, 7469–7473
29. Kohn, F., Zaltzman, P. A., Hager, M., and Itakura, K. (1984) Biochemistry 23, 7469–7473
MutS-Heteroduplex Interactions

32. Yang, C.-C., and Nash, H. A. (1989) Cell 57, 869–880
33. Lee, D. K., Horikoshi, M., and Roeder, R. G. (1991) Cell 67, 1241–1250
34. Jiricny, J., Hughes, M., Corman, N., and Rudkin, B. B. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 8860–8864
35. Lu, A.-L., Tsai-Wu, J.-J., and Cillo, J. (1995) J. Biol. Chem. 270, 23582–23588
36. Hunter, W. N. (1992) Methods Enzymol. 211, 221–251
37. Patel, D. J., and Shapiro, L. (1987) Annu. Rev. Biophys. Biophys. Chem. 16, 423–454
38. Kalnik, M. W., Norman, D. G., Swann, P. F., and Patel, D. J. (1989) J. Biol. Chem. 264, 3702–3712
39. Woodson, S. A., and Crothers, D. M. (1988) Biochemistry 27, 3130–3141
40. Rosen, M. A., Live, D., and Patel, D. J. (1992) Biochemistry 31, 4004–4014
41. Abou-la-ela, F., Murchie, A. I. H., Homans, S. W., and Lilley, D. M. J. (1993) J. Mol. Biol. 229, 173–188
42. Morden, K. M., Chu, Y. G., Martin, F. H., and Tinoco, I. (1983) Biochemistry 22, 5557–5563
43. van den Hoogen, Y. T., van Beuzekom, A. A., van den Elst, H., van der Marel, G. A., van Boom, J. H., and Altona, C. (1988) Nucleic Acids Res. 16, 2971–2986
44. Kalnik, M. W., Norman, D. G., Li, B. F., Swann, P. F., and Patel, D. J. (1990) J. Biol. Chem. 265, 636–647
45. Morden, K. M., Gunn, B. M., and Maskos, K. (1990) Biochemistry 29, 8835–8845
46. Rice, J. A., and Crothers, D. M. (1989) Biochemistry 28, 4512–4516
47. Wang, Y.-H., and Griffith, J. (1991) Biochemistry 30, 1358–1363
48. Hanawalt, P. C. (1995) Mutat. Res. 289, 7–15
49. Sancar, A. (1996) Annu. Rev. Biochem. 65, 43–81