MutS recognition: Multiple mismatches and sequence context effects

AMITA JOSHI and BASUTHKAR J RAO*

Department of Biological Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400 005, India

**Corresponding author (Fax, 91-22-2152110; Email, bjrao@tifr.res.in).*

Escherichia coli MutS is a versatile repair protein that specifically recognizes not only various types of mismatches but also single stranded loops of up to 4 nucleotides in length. Specific binding, followed by the next step of tracking the DNA helix that locates hemi-methylated sites, is regulated by the conformational state of the protein as a function of ATP binding/hydrolysis. Here, we study how various molecular determinants of a heteroduplex regulate mismatch recognition by MutS, the critical first step of mismatch repair. Using classical DNase I footprinting assays, we demonstrate that the hierarchy of MutS binding to various types of mismatches is identical whether the mismatches are present singly or in multiples. Moreover, this unique hierarchy is indifferent both to the differential level of DNA helical flexibility and to the unpaired status of the mismatched bases in a heteroduplex. Surprisingly, multiple mismatches exhibit reduced affinity of binding to MutS, compared to that of a similar single mismatch. Such a reduction in the affinity might be due to sequence context effects, which we established more directly by studying two identical single mismatches in an altered sequence background. A mismatch, upon simply being flipped at the same location, elicits changes in MutS specific contacts, thereby underscoring the importance of sequence context in modulating MutS binding to mismatches.

1. Introduction

The proteins that mediate the mismatch repair pathway are conserved across several species, from *Escherichia coli* to humans. As compared to most of the known repair proteins that recognize only a limited repertoire of DNA aberrations, MutS is more versatile since it can identify and bind to seven out of eight mismatches as well as loops that occur within the DNA (Su and Modrich 1986; Jiricny *et al* 1988; Parker and Marinus 1992; Modrich and Lahue 1996). Several eukaryotic homologues of *E. coli* MutS have been described, where different mismatches, insertion deletion loops (IDLs), etc. are recognized specifically by different sets of protein complexes (Modrich and Lahue 1996; Kolodner 1996; Kolodner and Marsischky 1999). Mismatch recognition is the critical initiating step in a series of events that culminate in post replication DNA repair and the proteins involved in it have evolved to specifically recognize features peculiar to mismatches *vis-a-vis* those that occur in normal Watson-Crick base pairs. A number of structural, thermodynamic and chemical features characterize mismatched base pairs and there has been sustained efforts in the field to determine the contribution of all these in invoking MutS recognition (Rajski *et al* 2000; Brown *et al* 2001).

Early attempts to associate the thermodynamic stability of mismatches with them being recognized and repaired by MutS suggested no clear correlation between the two (Werntges *et al* 1986). Inspite of extensive studies on the topic, the relationship between the helical flexibility *versus* different mismatches that cause it has also remained elusive (Lane and Peck 1995; Marathias *et al* 1999, 2000; Mol *et al* 1999). This may largely be due to the complex effects of sequence context on the helical flexibility and also due to the lack of reliable quantitative methods to measure the local as well as non-local aspects

Keywords. Mismatch; MutS; sequence context

Abbreviations used: DEPC, Diethyl pyrocarbonate; DTT, dithiothreitol; IDLs, insertion deletion loops.

of helix flexibility. A large body of structural data on both individual and closely spaced mismatches have revealed that most mismatches are stabilized by intrahelical hydrogen bonding as well as by base stacking interactions (Arnold *et al* 1987; Hunter *et al* 1987; Kouchaksjian *et al* 1988; Lane *et al* 1994; Gervais *et al* 1995; Boulard *et al* 1997; Allawi and SantaLucia 1998). A few mismatches do assume an extrahelical nature, but such effects were again found to be strongly sequence context dependent (Fazakerley *et al* 1986). Studies have also revealed that the fine structures of mismatched base pairs are not only context dependent but also extremely time dependent, i.e. dynamic in nature (Lane and Peck 1995; Patel *et al* 1984a,b,c). In light of this information, it was far from clear as to how a single protein such as MutS achieves specific recognition directed not only towards several conformational states of a single mismatch but also towards several different mismatches. One of the early attempts at trying to get an insight on this complex problem suggested that stacked, intrahelical base mismatches are better targets than the extrahelical ones for MutS binding (Fazakerley 1986). There has been extensive speculation about the features of mismatches that dictate repair enzyme specificity and the topic has been the subject of many reviews (Rajski *et al* 2000; Mol *et al* 1999; Jiricny 1998; Marra and Schar 1999).

Recent studies on the Taq MutS have revealed insights into the manner in which the protein makes intimate contacts, close to the mismatch, along both the major as well as the minor grooves (Biswas and Hsieh 1997; Malkov *et al* 1997). Recent high-resolution crystal structure data, for the first time, unveil some aspects of mismatch recognition principles of MutS (Obmolova *et al* 2000; Lamers *et al* 2000; Sixma 2001). These studies showed that a critical amino acid residue, phenylalanine, intercalates between a mismatched pair and a normal pair, facilitated by a large helical kink at the mismatch. A large shift of the T-base from its original major groove location to that of the minor groove has also been seen in GT-mismatch bound by MutS. All these aspects portend a complex recognition mode of MutS and demand further extensive studies on structural and dynamic aspects of this issue.

We have initiated studies to understand the basic "read out" principles that MutS might employ for recognizing mismatches. Our earlier studies demonstrated that MutS exhibits the most stable recognition mode towards a mismatch in the absence of any nucleotide cofactor, following which either ATP binding specific or hydrolysis specific conformational states ensue (Joshi *et al* 2000). With a view to elucidate the mechanisms of mismatch recognition by MutS, we describe experiments that analyse several molecular determinants of the DNA helix that seem to be significant in this process. Firstly, neither the global flexibility of the helix nor the local pairing status of the mismatches seems to be directly relevant in invoking MutS binding. Interestingly, sequence context plays a major role that is manifested by changes in MutS contacts on a GT-mismatch following a simple flipping of the same.

2. Materials and methods

2.1 *Materials*

Oligonucleotides were synthesized at DNA Technology (Denmark) or at the Keck Biotechnology Resource Laboratory (Yale, USA). T4 polynucleotide kinase, DNase I, dithiothreitol (DTT) and diethyl pyrocarbonate (DEPC) were from Amersham Life Science. Nuclease-free BSA was from Sigma. $KMnO_4$ was obtained from Merck.

2.2 *DNA substrates*

Table 1 lists all the substrates used in this study. The purification of oligonucleotides was carried out as described (Joshi *et al* 2000). Unless otherwise mentioned, DNA concentrations expressed refer to molar concentrations of oligonucleotide molecules.

2.3 *End labelling of oligonucleotides*

Oligonucleotides were end labelled using T4 polynucleotide kinase and [g-^{32}P] ATP as described (Joshi *et al* 2000).

2.4 *Annealing of strands*

Annealing (for all 33 bp duplexes) was done by mixing the labelled strand (1·2 μM) with unlabelled strand (1·8 μM) in a total volume of 10 μl followed by heating at 90°C for 4 min and slow cooling to room temperature in 20 mM Tris Cl (pH 7·6) and 5 mM $MgCl_2$. Completion of annealing was assessed by analysis on native polyacrylamide gel, which showed that more than 90% labelled strand was converted to duplexes and no residual unannealed labelled single-stranded DNA was present. The annealed product was not an artifact of self-annealing of labelled strand, as mobility of appropriate self-annealing controls were compared in the same native polyacrylamide gel in each case. At these DNA concentrations and annealing conditions, the presence of multiple mismatches between the complementary strands did not seem to significantly affect the extent of heteroduplex annealing. Therefore, all the experiments involving labelled DNA represent binding of MutS to the duplex form of DNA.

2.5 *Protein purification*

Purification of the protein was carried out as described (Joshi *et al* 2000; Worth *et al* 1998).

2.6 *MutS binding to heteroduplexes*

Labelled duplexes (0·03 μM, unless otherwise mentioned) were incubated with MutS (2·5 μM, unless otherwise mentioned) in a binding buffer containing 20 mM Tris Cl (pH 7·6), 50 mM KCl, 5 mM $MgCl_2$, 2 mM DTT, 50 μg/ml nuclease free BSA, and 0·4 μM oligo dT (30 mer) as carrier DNA in a final volume of 10 μl at 0°C for 30 min (for gel shift assays), or 37°C for 15 min (for DNase I footprinting assays).

2.7 *DNase I footprinting assays*

DNase I footprinting assays were carried out as described (Joshi *et al* 2000).

2.8 *Chemical modification experiments*

2.8a *DEPC reactions*: Substrates (K, L, M, and the WT homoduplex, see table 1) (0·03 μM) were treated with DEPC (4%) in the presence of 20 mM Hepes-NaOH (pH 7·6), 5 mM $MgCl_2$, 1 mM EDTA, and 1 mg/ml salmon sperm DNA at 20°C for 20 min. DEPC reaction was quenched by the addition of carrier t-RNA (250 μg/ml) that acts as excess target substrate, followed quickly by the addition of chilled ethanol (75%) that precipitates the DNA, thereby squelching the DNA away from DEPC reagent in solution, whose reactivity is already lowered

Table 1. DNA-substrates.

Substrate	Strand	Sequence
A (TG)	At	5′ CAT GGA GCA G**T**T CGC GGA TTT CGA CAC AAT TTA 3′
A	Ab	3′ GTA CCT CGT C**G**A GCG CCT AAA GCT GTG TTA AAT 5′
B (GT)	Bt	5′ CAT GGA GCA G**G**T CGC GGA TTT CGA CAC AAT TTA 3′
B	Bb	3′ GTA CCT CGT C**T**A GCG CCT AAA GCT GTG TTA AAT 3′
D, E, F	DEFt	5′ GGC TTA GAG CT**T** AAT TGC TGA A**T**C TGG TGC TGT 3′
D (TG)	Db	3′ CCG AAT CTC GA**G** TTA ACG ACT T**G**G ACC ACG ACA 5′
E (TG)	Eb	3′ CCG AAT CTC GAA TTA ACG ACT T**G**G ACC ACG ACA 5′
F (TG)	Fb	3′ CCG AAT CTC GA**G** TTA ACG ACT TAG ACC ACG ACA 5′
	MMt	5′ TAA ATT GTG TCG AAA TCC GCG ACC TGC TCC ATG 3′
G (GT)	Gb	3′ ATT TAA CAC AGC TTT AGG **T**GC TGG ACG AGG TAC 5′
WT	Hob	3′ ATT TAA CAC AGC TTT AGG CGC TGG ACG AGG TAC 5′
H (GT)	Hb	3′ ATT TAA **T**A**T** AG**T** TTT AGG **T**G**T** TGG A**T**G AGG TA**T** 5′
I (GG)	Ib	3′ ATT TAA **G**A**G** AG**G** TTT AGG **G**G**G** TGG A**G**G AGG TA**G** 5′
J (GA)	Jb	3′ ATT TAA **A**A**A** AG**A** TTT AGG **A**G**A** TGG AA**G** AGG TA**A** 5′
K (AC)	Kb	3′ A**C**C **C**AA CAC AGC **C**C**C** AGG CGC **C**GG ACG AGG **C**AC 5′
L (AA)	Lb	3′ A**A**A **A**AA CAC AGC **A**A**A** AGG CGC **A**GG ACG AGG **A**AC 5′
M (AG)	Mb	3′ A**G**G **G**AA CAC AGC **G**G**G** AGG CGC **G**GG ACG AGG **G**AC 5′
N (TC)	Nb	3′ **C**TT T**C**C C**C**C **C**GC TTT **C**GG CGC TGG **C**CG **C**GG T**C**C 5′
O (TG)	Ob	3′ **G**TT T**G**G C**G**C **G**GC TTT **G**GG CGC TGG **G**CG **G**GG T**G**G 5′
P (TT)	Pb	3′ **T**TT T**T**T C**T**C **T**GC TTT **T**GG CGC TGG **T**CG **T**GG T**T**C 5′
Q (CA)	Qb	3′ ATT TAA CAC A**A**C TTT A**A**A C**A**C T**A**A ACA AAA TAC 5′
R (CT)	Rb	3′ ATT TAA CAC A**T**C TTT A**T**T C**T**C T**T**T ACT A**T**T TAC 5′
S (CC)	Sb	3′ ATT TAA CAC A**C**C TTT A**C**C C**C**C T**C**C AC**C** A**C**C TAC 5′

't' and 'b' suffix in the strand column refers to 'top' and 'bottom' strands respectively of an annealed duplex.
'DEFt' is the common 'Top' strand that upon annealing with Db, Eb, and Fb generates duplexes D, E, and F respectively.
'MMt' is the common 'Top' strand that upon annealing with Gb, Hob, Hb, Ib, Jb, Kb, Lb, Mb, Nb, Ob, Pb, Qb, Rb, and Sb generates duplexes G, WT, H, I, J, K, L, M, N, O, P, Q, R and S respectively. Duplex substrates thus generated are listed in the first column along with the mismatches that they contain in parenthesis. Mismatches are shown in bold letters. The first letter of any mismatch represents a base present in the top strand and the second letter in the bottom strand.
All 'Top' strand sequences are in 5′–3′.
All 'Bottom' strand sequences are in 3′–5′.
WT represents wild-type sequences that contain no mismatches (homoduplex).

significantly by the drop in solution temperature. The precipitate was washed twice with 70% ethanol, followed by drying in a speed vac. Piperidine (70 µl, 10%) cleavage was done by incubating at 90°C for 30 min. The samples were dried and washed extensively with water (thrice) to remove all traces of residual piperidine. The dried pellets were dissolved in 50% formamide containing bromophenol blue and analysed on a 12% denaturing gel.

2.8b *$KMnO_4$ reactions*: Substrates (O, N, P, and the WT homoduplex, see table 1) (0·03 µM) were treated with $KMnO_4$ (6 mM) at 37°C for 2 min in the standard binding buffer in 10 µl reactions. The reactions were quenched by the addition of 3 µl of standard DMS stop solution (1·5 M Sodium acetate, 1 M *b*-mercaptoethanol, 250 µg/ml t-RNA). The subsequent treatment of the samples was carried out in the same manner as that for the DEPC reactions.

3. Results

In this study, we have focused on parameters that influence the overall physicochemical properties of duplex DNA containing mismatches, namely, the sequence context effects and backbone flexibility. In the present study, we have employed various biochemical (DNase I and chemical footprinting) approaches to assess the role of these components in modulating the recognition of mismatches by MutS. The heteroduplexes studied contained either single/double or multiple mismatches (seven to nine) in a 33-mer duplex. It is pertinent to mention here that this is the first biochemical study of MutS on multiple mismatches, the motivation for which was two-fold: (i) Multiple mismatches induce enormous flexibility even in very short duplexes that are far shorter than the persistent length of a duplex and offer an excellent opportunity to address the issues related to helical flexibility *versus* MutS recognition of a given mismatch. (ii) Classical genetic experiments have demonstrated that MutS is critically involved in the recognition and the subsequent abrogation of multiple mismatches that are generated during homologous recombination between genomes that are highly divergent (Rayssiguier *et al* 1989). We wanted to learn the rules of MutS recognition hierarchy that may be relevant for its editing function in such recombination reactions.

3.1 *Recognition hierarchy of mismatches by MutS is same whether they occur as single or multiple*

As described in the table 1, we substituted either all A's, T's, G's or C's of a 33-mer strand (Hob, table 1) with any other base followed by annealing each one to a common strand (MMt), thus generating heteroduplexes (substrates H-S) where 25% of the base pairs are mismatches of a specific type (Karthikeyan *et al* 1998). Amongst the twelve heteroduplexes so designed, a set of three heteroduplexes, which contained AC/AG/AA or GT/GA/GG or CA/CT/CC or TG/TC/TT mismatches have identical sequence background and hence form a set for comparison. Since all the heteroduplexes contain large number of mismatches, we made sure on native polyacrylamide gels that the labelled products under study are truly duplex products largely free of residual unannealed labelled single strands (see §2). We monitored the binding of MutS to all twelve heteroduplexes using a *quantitative* gel-shift assay. Specific gel-shifted complexes were *quantified* (figure 1A). Under the same conditions of electrophoresis, no shifts were observed in controls containing either homoduplexes or single-strands. A clear hierarchy of MutS binding to mismatches was observed which was identical to that reported for single mismatch containing substrates (Su *et al* 1988). As reported earlier, the mismatches TG, GT, CA and AC were best recognized by MutS, followed by a moderate binding to GG and AA (figure 1A). MutS poorly recognized the remaining mismatches. MutS binding to the same substrates was also monitored by DNase I footprinting, an assay that monitors a reaction at equilibrium better as compared to the gel-shift assay. Nine out of the twelve heteroduplexes were footprinted using a labelled common strand (MMt in table 1). In all the three comparable sets, namely, GT, GG and GA; AC, AA and AG; TG, TT and TC, DNase I protection due to MutS binding mirrored the gel-shift assay results (figure 1B). For example, a decreasing order of protection as well as binding was observed across GT, GG and GA mismatches. The same trend was observed across sets AC, AA and AG and TG, TT and TC. These two experiments, taken together, revealed that multiple mismatch substrates seem to retain local elements of structure specific to each mismatch, in spite of the high percentage of mismatches in the duplex. Although the hierarchy of recognition by MutS remains the same in multiple *versus* single mismatches, whether the intrinsic affinity towards the protein was altered was not known. To test this we compared MutS binding to single *versus* multiple GT mismatches in duplexes with identical sequence background. In the parental strand (Hob), either a single C in the middle (strand Gb) or all the C's across the strand (strand Hb) were substituted by T, followed by annealing to a common complementary strand (MMt). Annealing in the former substrate generates a centrally located GT mismatch, whereas in the latter the mismatches are distributed along the duplex. Therefore, the single GT mismatch is one among the multiple GT mismatches in an otherwise identical sequence background. MutS titration with either substrates followed by DNase I footprinting revealed substantial differences in

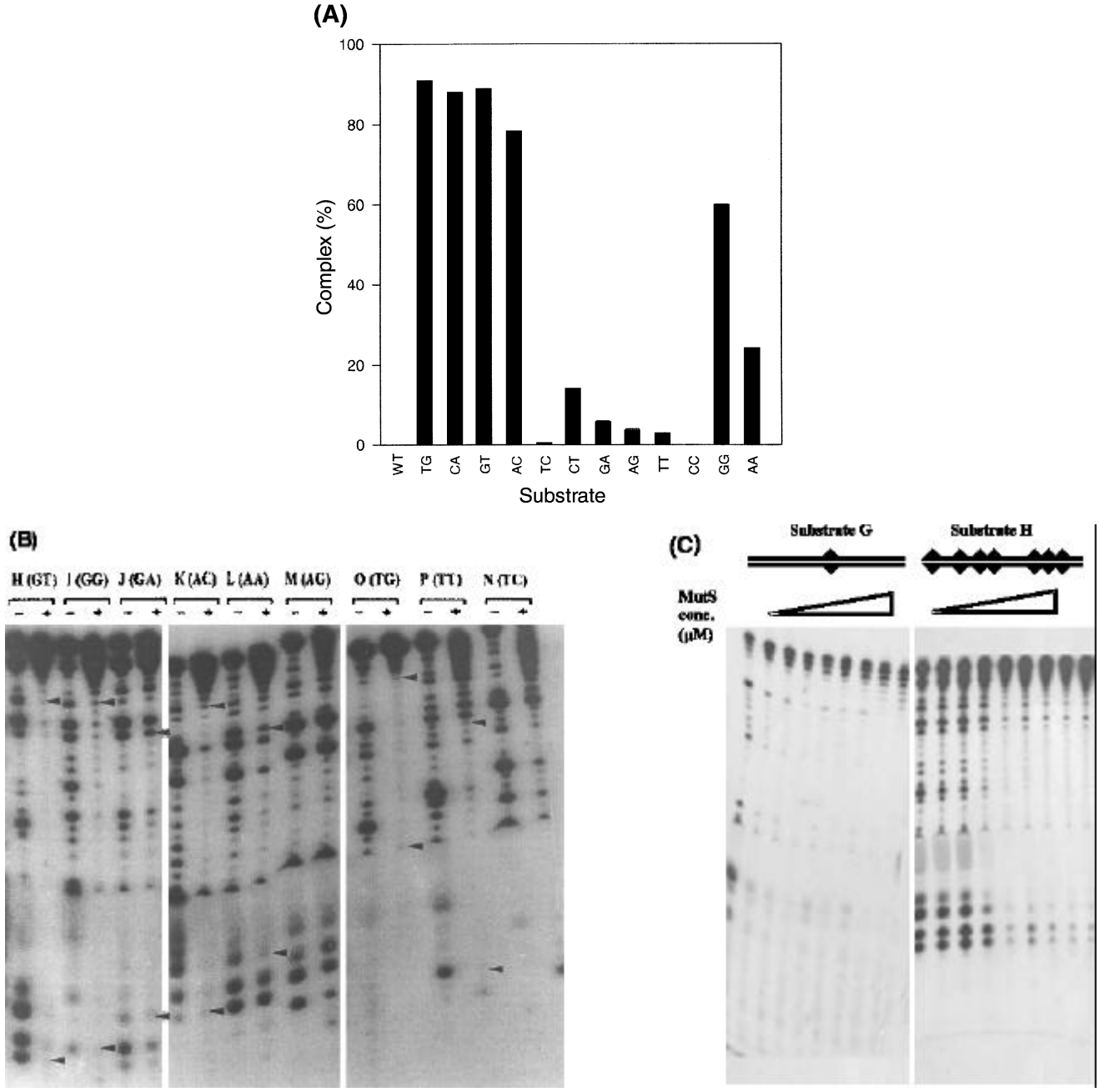

Figure 1. Hierarchy of MutS binding to various multiple mismatches: gel-shift assay (**A**) and DNase I protection assay (**B**). Twelve different heteroduplexes (H to S) were generated by annealing the labelled common top strand (MMt) with any one of the bottom strands (Hb to Sb) (table 1). Control homoduplex (WT) was generated by annealing the labelled strand (MMt) with the unlabelled complementary strand (Hob). MutS binding reaction was carried out at the standard conditions (see §2) followed by gel-shift analysis on a precooled, native 6% polyacrylamide gel (Joshi *et al* 2000). MutS-mismatch-DNA, which revealed as gel-shifted complexes, were quantified by PhosphorImager analysis and expressed as percentage radioactivity of total annealed duplex. In the depiction of a mismatched base-pair, first base is carried by the common labelled top strand (MMt) whereas the second one is carried by any of the unlabelled strands from a set (Hb to Sb). For DNase I protection assays, all substrates except Q, R and S were used. MutS binding followed by DNase I protection was done as described previously (Joshi *et al* 2000). The letters on the top of each lane denote the substrate and the corresponding mismatch, whereas '–' and '+' indicate the absence and presence of MutS respectively. Arrowheads mark the boundaries of DNase I protection. (**C**) Affinity of MutS to single GT *versus* multiple GT mismatches. Substrates G and H, representing single and multiple GT mismatch containing duplexes respectively, were formed by annealing strands Gb and Hb with a common labelled strand (MMt). The labelled duplexes were incubated with increasing concentrations of MutS (0, 0·25, 0·5, 0·75, 1·0, 1·5, 2·0, 2·5, and 5·0 μM, in each set) followed by DNase I footprinting assay. The relative locations of the mismatches are schematically represented by filled triangles on the line diagrams of the duplexes.

the affinity of MutS (figure 1C). Surprisingly, the heteroduplex containing multiple mismatches (substrate H) showed DNase I protection only at high concentrations of MutS (1 μM), as against single GT mismatch heteroduplex (substrate G) which was protected by 0·25 μM MutS.

These results suggest that individual mismatches within a multiple mismatch duplex maintain sufficiently specific local features of base mispair chemistry due to which hierarchy of recognition of multiple mismatches by MutS is same as that for single mismatches. The stability of the MutS-mismatch complex, which is also governed by the neighbouring sequences to maintain a normal Watson-Crick helix, is probably compromised in the multiple mismatch heteroduplex, thereby resulting in reduced binding affinity for MutS.

3.2 *Lack of correlation between helix flexibility and MutS binding affinity*

When studied as a naked polymer in solution, DNA has limited flexibility. DNA stiffness is reflected in estimates of its persistence length (about 140 bp) (Hagerman 1988). What are the structural consequences of mismatches within a duplex and what is the relationship, if any, between the number of mismatches and such consequences? The structures of mismatched base pairs within short duplexes have been investigated by high resolution techniques such as NMR and X-ray crystallography. Several single mismatches have been shown to stack into the helix with minimal helical distortion (Hunter *et al* 1987; Lane *et al* 1994; Gervais *et al* 1995; Allawi and SantaLucia 1998; Bhattacharya and Lilley 1989a,b). On the other hand, depending upon the sequence context, some bases do flip out and assume an extrahelical configuration (Fazakerley *et al* 1986). There is no clear insight about the relationship between such structural aberrations due to mismatches *versus* their recognition by MutS. The multiple mismatch system, that retains the hierarchy of MutS binding similar to that observed for single mismatches, offers an opportunity to address some aspects of this important relationship.

Gel electrophoresis is very sensitive to alterations in the path of helical axis and can readily detect trajectory changes due to sequence-directed curvature and protein-induced bending (Wu and Crothers 1984). Elegant studies that employed gel electrophoretic migration, helical phasing experiments as well as DNA cyclization kinetics have demonstrated that clustered Pur-Pur and Pyr-Pyr mismatches impart localized flexibility in DNA helix (Kahn *et al* 1994). We wanted to analyse whether multiple mismatch containing duplexes studied here show any such macroscopic distortions in the helix that can be detected by electrophoretic migration assays. All mismatched duplexes of three comparable sets, namely, GT, GG and GA; AC, AA and AG; TG, TT and TC were analyzed. Duplexes of the first set, namely, GT, GG and GA migrated at the same rate as a homoduplex control of the same size (figure 2A). On the other hand, AC, AA and AG revealed differences in mobility. All the three were slower than the homoduplex control and AG was the fastest in the set. Amongst TG, TT and TC, the differences were greater as the migration rate increased from TT < TG < TC and TC comigrated with the homoduplex control. The slower mobility of some of these heteroduplexes is a reflection of a higher degree of flexibility inherent within the helix than that in the corresponding homoduplex. It is important to mention here that the radioactive bands depicted in the figure belong to the annealed duplexes only. The left over labelled single-strands are not shown. Moreover, the faint signals associated with AC and TT lanes were due to inadvertent under-loading of the two samples in this experiment. Repeat analyses of the same samples confirmed that the faint signals did indeed represent the mobility positions of genuine AC and TT duplexes (data not shown).

It was also interesting to know whether the mismatch imparted flexibility in the duplex influenced the extent of their recognition by MutS. For example, within a set (TG, TT, TC) the TT heteroduplex, which exhibits the highest backbone flexibility as revealed by its slowest migration rate, is very poorly bound by MutS (figure 1A). In contrast, the protein also poorly recognizes TC whose migration rate is the same as that of the homoduplex. On the other hand, TG with an intermediate flexibility is recognized best by MutS. The flexibility of GT, GG and GA was similar, even though MutS binding ability to these mismatches was largely dissimilar (figure 1A). These results clearly suggested that there was no obvious link between the parameter of helical flexibility and MutS affinity towards them (see §4).

3.3 *The unpaired status of a mismatch may not invoke recognition by MutS*

We wanted to examine whether local parameters specifically related to the mismatch as against more global features such as that described above (duplex flexibility) play a significant role in invoking mismatch recognition. Mismatches, as against normal Watson-Crick base pairs, exhibit varying degrees of stability. This stability is a function of the pairing status of the mismatch, which might involve “wobbling” of mismatched base pairs as a result of a high rate of interconversion between different states of pairing and lead to base pair “opening” (Lane and Peck 1995; Patel *et al* 1984a,b,c; Plum and Breslauer

1994). We addressed whether the "wobble" nature of mismatched base pairs leads to a single stranded state/ unpairing at the site of the mismatch at equilibrium and consequently invokes recognition by MutS. This assumption also stemmed from the findings that the GT mispair which exhibits considerable degree of "wobble" structure (Allawi and SantaLucia 1998) is best recognized by MutS, whereas AG, which is one of the most stably paired mismatches is least recognized. We tested this hypothesis by carrying out chemical footprinting on sets of multiple mismatched substrates. DEPC is known to bring about carbethoxylation of adenine bases at the N7 position, and the extent of this modification is higher at sites where adenines are unpaired (Bhattacharya and Lilley 1989a, b; Leonard *et al* 1971; Vincze *et al* 1973). Substrates, which bear either multiple AC/AA/AG mismatches at identical

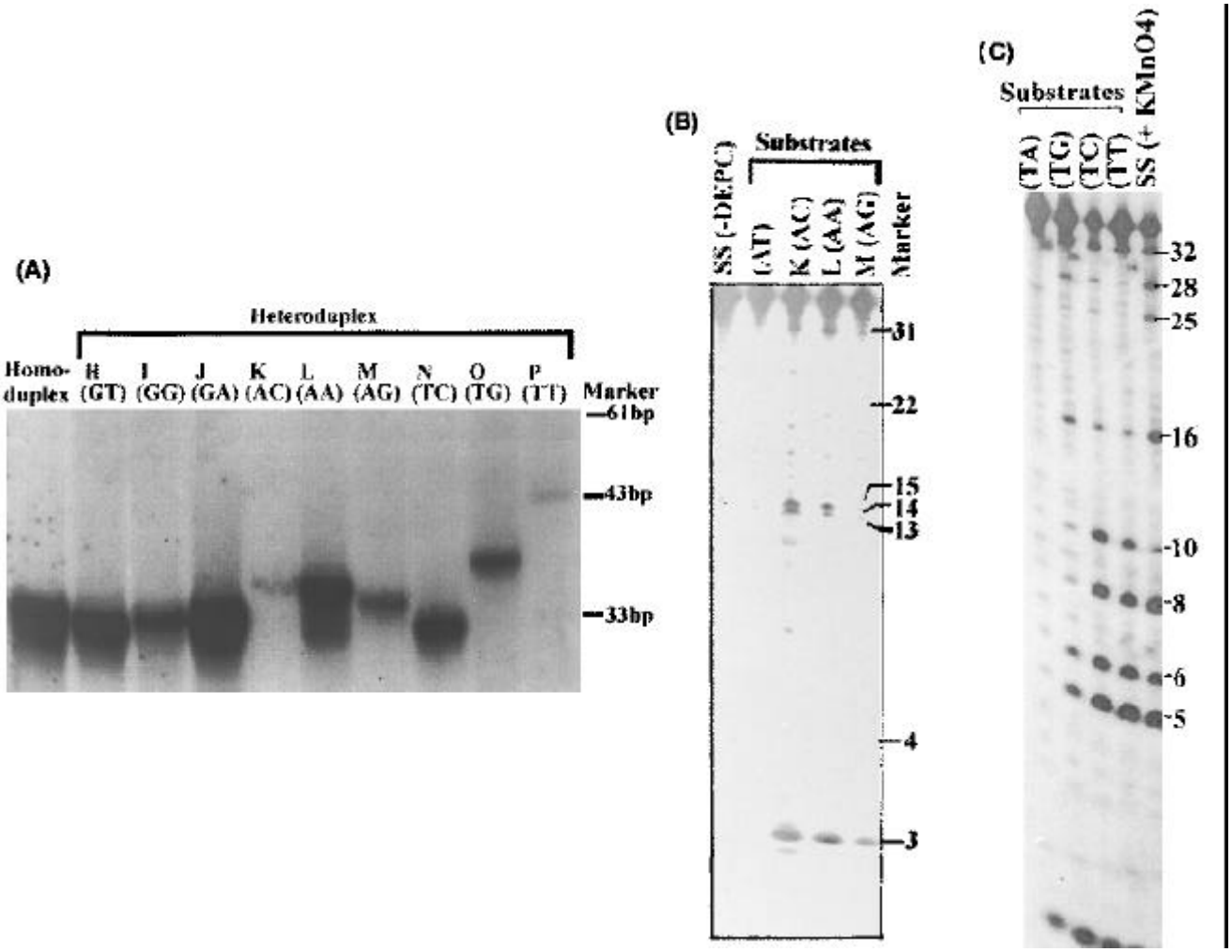

Figure 2. (**A**) Native gel electrophoretic analysis of helical flexibility of heteroduplexes containing multiple mismatches. The duplex substrates used were prepared as described in the legend for figure 1A. The annealed samples were electrophoresed on a precooled 12% native polyacrylamide gel (500 V, 25 × 40 × 0·2 cm gel) at 4°C for 6 h. The gels were then dried and autoradiographed. The positions of molecular weight markers are as indicated. (**B**) DEPC probing of A residues in multiple AC, AA, AG and AT pairings. Substrates K, L, M and homoduplex were prepared by annealing the labelled common strand (MMt) with the respective unlabelled bottom strands (Kb, Lb, Mb, and Hob) (see §2). The substrates (0·06 μM) were subjected to DEPC modification, followed by piperidine cleavage (see §2). The lane SS (-DEPC) indicates piperidine cleaved labelled control single strand (MMt) that monitors the background cleavage. The numbers on the right (Marker) represent the positions of the corresponding adenine cleavage products. (**C**) $KMnO_4$ probing of T residues in multiple TG, TC, TT and TA pairings. Substrates O, N, P and homoduplex were prepared by annealing the labelled common strand (MMt) with the respective unlabelled bottom strands (Ob, Nb, Pb and Hob) (see §2). The substrates (0,06 μM) were subjected to $KMnO_4$ modification, followed by piperidine cleavage (see §2). The lane SS (+ $KMnO_4$) indicates the modification of the labelled single strand (MMt). The numbers on the right indicate the positions of the thymine cleavage products.

positions in an otherwise same sequence background, were probed. In all the duplexes, the common top strand (MMt) was labelled. Homoduplex control that contained the same top strand labelled showed no DEPC modification at any of its adenine residues (figure 2B). We observed a high degree of non-uniformity in the sensitivity of various adenine residues across all mismatches in a given heteroduplex. For example, the adenine clusters at locations 2–4 and 13–15 showed reactivity to DEPC, whereas those located singly at positions 31 and 22 did not. Even within a cluster, not all adenines were equally reactive. The reactivity appears to be a complex function of sequence context effects. Moreover, for reasons unclear to us, two pyrimidine residues on either side of the 31st adenine showed reactivity to DEPC. The degree of bonafide DEPC modifications observed at adenine clusters 2–4 and 13–15 were different across the three heteroduplexes studied. For example, AC mismatches showed the highest modification followed by AA and AG. In fact, AG mismatches barely showed any modification for adenines in the cluster 13–15. Since the degree of DEPC modification is a reflection of the unpaired status of the adenine residues, the gradation in DEPC modification observed here suggests that unpairing of adenine residues follows a pattern where AC > AA > AG. This correlated well with the hierarchy of MutS binding for AC, AA and AG (figure 1A,B). However, the affinity differences shown by MutS towards AC, AA and AG heteroduplexes, cannot be reconciled with the uniform pairing seen at the 22nd and 31st mismatches across all three heteroduplexes. Such a discrepancy becomes more obvious in the next chemical modification experiment.

We used $KMnO_4$ reactivity to monitor the relationship between MutS binding and the status of base "unpairing" in the mismatches using TC, TG and TT heteroduplex set. Unlike the adenines in the DEPC experiment, most thymines in heteroduplexes showed reactivity towards $KMnO_4$ (figure 2C). Moreover, the optimized conditions of $KMnO_4$ modification employed in this experiment revealed a high degree of specificity towards all the thymine residues only when they were present in the single stranded form, as shown in a single stranded DNA control (common labelled top strand, MMt, table 1). In contrast, the thymines in the homoduplex control barely showed any reactivity. For the thymines in heteroduplexes, we observed three categories of reactivity towards $KMnO_4$ depending upon the nature of the mismatch. Thymines at the 32nd position were uniformly reactive in TG, TC and TT heteroduplexes. On the other hand, those at the 28th and 16th positions showed a decreasing order of reactivity in TG, TC and TT heteroduplexes, while the trend was exactly reversed for the mismatches at positions 10, 8, 6, 5 and 1. However, the binding hierarchy observed in this set had shown that TG was the best binder followed by TT and TC (figure 1A,B). Thus, this experiment strongly suggested a complete lack of correlation between the unpaired status of thymines in heteroduplexes and their binding to MutS. The findings from the DEPC and $KMnO_4$ footprinting experiments, taken together, revealed an absence of any obvious association between the unpaired status of a mismatch and its ability to bind MutS (see §4).

3.4 *MutS affinity towards mismatches is governed by sequence context*

A given mismatch can occur in the DNA in multiple sequence backgrounds. To know if the intrinsic affinity of MutS towards the same mismatch in two different sequence backgrounds is the same, or does it show any significant differences, we positioned two TG mismatches in the same duplex within a short distance of each other, such that the binding of MutS to one mismatch might sterically interfere with the binding at the other. We analysed whether MutS makes a choice between the two mismatches in a "selective" or a "stochastic" binding mode. The former outcome would support the importance of sequence context in modulating MutS affinity for a mismatch, whereas, the latter would negate the same. A heteroduplex containing two TG mismatches was compared with those that contained either of the two mismatches singly, in a DNase I footprinting assay (Substrate D *versus* substrates E and F, figure 3A). The footprint positions on single mismatch substrates were distinctly different from each other on top-strands, but not so on the bottom-strands. Moreover, both footprints encompassed about 12–14 nucleotides each. However, the footprint on double mismatched substrate (substrate D) showed a bias towards one of the mismatches (that occurring in substrate E) with respect to the top-strand (figure 3A). The bottom-strand protection encompassed essentially the same region and was common for either of the single mismatch. Therefore, the bias observed in the top-strand footprint of the double-mismatch substrate seems to suggest a selective rather than a stochastic mode of mismatch recognition. Amongst the two identical mismatches on the same duplex, MutS binding was preferential towards one, which highlighted the influence of sequence context on MutS recognition of a mismatch.

To determine whether is such a bias between two similar mismatches is a simple outcome of differences in the binding affinity of MutS, we assessed the same by a protein titration experiment followed by footprint analyses. The TG mismatch in substrate E started showing a footprint at the first concentration of MutS itself (0·25 μM) and yielded a stable footprint by 0·5 μM, whereas the same in substrate F did so only at as high a concentration

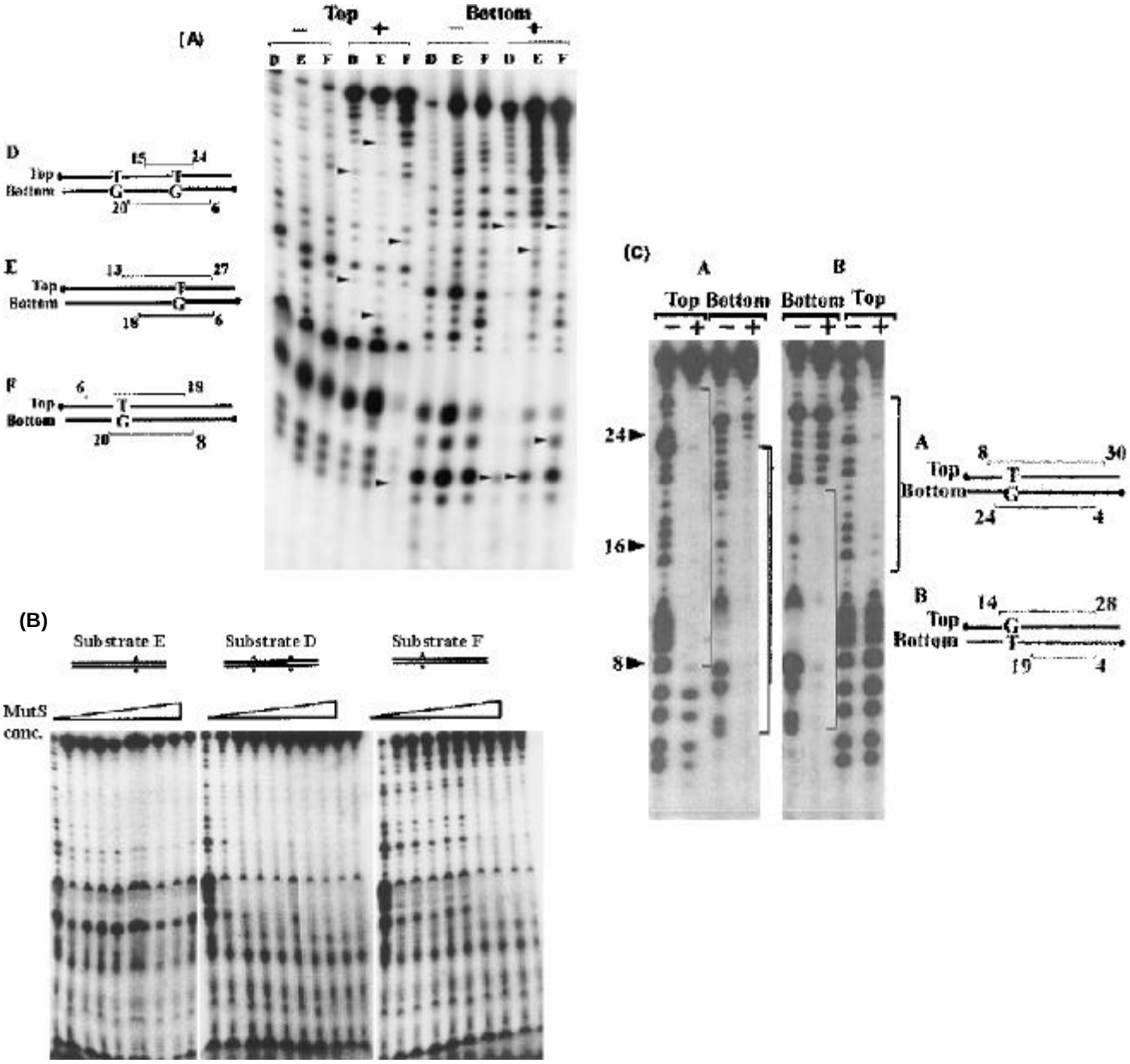

Figure 3. (**A**) DNase I footprinting of two separate TG mismatches on the same duplex: analysis of sequence context effects. The substrates (D, E and F) were prepared by annealing the labelled common top strand (DEFt) with the respective unlabelled bottom strands (Db, Eb and Fb) (first six lanes). The same substrates, with the bottom strand labelled (last six lanes), were prepared by annealing unlabelled common top strand (DEFt) with the respective labelled bottom strands (Db, Eb and Fb). Standard conditions were used for MutS binding and DNase I footprinting (see §2) (Joshi *et al* 2000). The boundaries of the footprints are indicated by arrowheads in each case. '+' and '–' indicate the presence and absence of MutS respectively. A schematic representation of the positions of mismatches on the three substrates and the relative nucleotide locations of the footprints on each are given. (**B**) Assessment of the relative affinities of MutS for two different TG mismatches present on the same duplex. The substrates (D, E and F) with the top strand (DEFt) (figure 3A legend) labelled were incubated with increasing concentration of MutS (0, 0·25, 0·5, 0·75, 1, 1·5, 2·0, 2·5, 5·0 and 7·5 μM, in each set) followed by DNase I footprinting assay. The filled triangles in the schematic duplex represents the relative locations of the TG mismatches. (**C**) Flipping of a TG mismatch into a GT mismatch alters MutS footprints. Substrate A was generated by annealing strands At and Ab, where either the former (first two lanes) or the latter (next two lanes) was labelled. Similarly, substrate B was formed by annealing strands Bt and Bb, where either the former (last two lanes) or the latter (first two lanes) was labelled. Substrate A and B contained a TG and a GT mismatch respectively at the same position. '+' and '–' indicate the presence and absence of MutS respectively. The footprinted regions are bracketed. The numbers with the arrowheads on the left represent the nucleotide positions. A schematic representation of the relative nucleotide locations of the footprints on each is given.

of MutS as 2·0 μM (figure 3B). The same mismatch, in two different sequence contexts, showed an approximately ten-fold difference in MutS affinity. In the same experiment, MutS titration with the double mismatch substrate yielded a stable footprint by 0·5 μM MutS, just as that observed with substrate E that contained a TG mismatch of higher affinity. The result with the double mismatch substrate revealed that the non-stochastic mode of mismatch selection by MutS stems from the intrinsic affinity differences arising due to sequence context effects. If the sequence context effects are important, a mismatch in a heteroduplex would appear significantly different to MutS when the mismatch is simply flipped. It is only in the context of the surrounding sequences that a mismatch can be differentiated from its flipped version. To study the same, we footprinted the MutS binding on a TG mismatch (substrate A) and compared it with that of a GT mismatch located at the same site (flipped version of the TG mismatch, substrate B), either of which were known to be recognized by MutS equally well (figure 1A). Flipping of the mismatch resulted in a distinct alteration of the MutS-heteroduplex footprint on either strand (figure 3C). The footprint reduced in size by about 8 nucleotides and 5 nucleotides on the top and bottom strands respectively, simply due to flipping of TG to GT in an otherwise identical sequence background. The footprint that encompassed either side of the mismatch (TG) shifted almost entirely to one side of the mismatch (GT) following flipping. This experiment revealed that at the same site in a heteroduplex, TG *versus* GT is read differently by MutS owing to the contextual influence of sequences surrounding the mismatch.

4. Discussion

The experiments described in the present study reveal new insights about parameters that influence the specific recognition of a mismatch by MutS. We first examined the role of a global feature of mismatch containing duplexes, i.e. helical flexibility, in recruiting MutS to mismatches. Since single mismatches are known to cause minimal change in the helix backbone, we used a multiple mismatch system where introduction of high numbers of mismatches in a duplex significantly changes the helix flexibility thereby allowing an assessment of this parameter in MutS binding. Since MutS, due to its anti-recombination function, is known to recognize and abrogate multiple mismatches in a heteroduplex (Rayssiguier *et al* 1989), this study was intended to get insights on the nature of multiple mismatch recognition by MutS. Firstly, the experiments with multiple mismatches suggest that global changes in the helix (helical flexibility) do not seem to critically influence the overall binding hierarchy of mismatches by MutS (figure 2A). Electrophoretic mobility assays devised by Khan *et al* (1994) had elegantly demonstrated that a cluster of contiguously placed Pur-Pur or Pyr-Pyr mismatches substantially increased the torsional and bending flexibility of the helix. Using this assay, we showed that short heteroduplexes containing multiple mismatches exhibit different levels of helical flexibility depending upon the types of mismatches they contain. Under the same gel electrophoretic conditions, none of the single mismatches showed any significant mobility retardation with respect to the homoduplex controls (data not given). This result is consistent with reported observations on single *versus* multiple mismatched DNA substrates, where it was observed that while single mismatches do not significantly alter the path of the helical axis, the multiple mismatches (bubbles) lead to flexible helices (Bhattacharya and Lilley 1989a,b). Interestingly, a direct comparison of MutS binding strength towards a given type of mismatch *versus* the relative helical flexibility of the duplex housing that mismatch revealed no correlation between the two (figures 1A,B and 2A).

Some mismatches have been known to exhibit a "wobble" structure, which often lead to "base pair opening". In light of previous findings, which suggest better recognition of unpaired mismatches by MutS (Allawi and SantaLucia 1998; Fazakerley *et al* 1986) we addressed the issue by using a chemical footprinting approach. On probing the pairing status of several mismatches we observed that bases in different mismatches show different degrees of "unpairing" depending upon their pairing partner as well as the sequence in which they lie embedded. In a set of substrates where the sequence background was kept identical, the unpairing status of specific sets of A amongst AC, AA and AG, followed a decreasing level in that order (figure 2B). Similarly, in another set of comparison (TG, TC and TT), the unpaired status of T was non-uniform not only across different types of mismatches, but also between the same mismatches at different locations in the same heteroduplex (figure 2C). A direct comparison of MutS binding and the unpaired nature of mismatches in these sets failed to reveal any simple correlation (figures 1A,B, and 2B,C). At present, not much is known about the conformational dynamics associated with various mismatches either in single mismatch or in multiple mismatch settings. One of the studies has clearly indicated at the possibility of rapid interchange between different conformational states of GG mismatches in various heteroduplexes (Lane and Peck 1995). In view of this, it is reasonable to speculate that MutS might selectively recognize only a fraction of the conformational states of a mismatch and thereby shifting the dynamic equilibrium intrinsic to a mismatch towards protein bound form. Therefore, neither the macroscopic (helical flexibility) nor the local properties (unpaired

status of base mismatches) of the DNA helix seem to correlate in any way with aspects of recognition by MutS.

The hierarchy of MutS binding to various mismatches was similar irrespective of whether the heteroduplex contained single or multiple mismatches of a specific type (figure 1A,B). This seems to suggest that elements of local chemistry associated with a mismatch can override the non-local, macroscopic elements of duplex DNA structure in multiple mismatch containing helices. Elsewhere, we have showed more directly that a single subtle change in the chemical functional group of a Watson-Crick base pair chemistry can "trick" MutS into a specific recognition of the same (unpublished results). Recent studies by others also reveal the critical importance of local features of chemistry that provoke specific recognition by MutS (Brown *et al* 2001). The result that multiple mismatches cause significant reduction in MutS affinity towards them is entirely counter-intuitive (figure 1C). This might relate to the overall effects of sequence context which regulate not only the affinity of MutS towards a mismatch (figure 3B) but also the fine details of the way MutS contacts a mismatch *vis-a-vis* the surrounding sequences (figure 3C). Although studies in the past have documented that the surrounding sequence of a mismatch influences MutS binding both *in vitro* as well as *in vivo*, there is no clear insight about the molecular determinants that mediate these effects (Fazakerley *et al* 1986; Jones *et al* 1987; Brown *et al* 2001). In our studies, we see clear effects of sequence-context on binding affinity, as well as the type of physical contacts made between MutS and a mismatch.

It is interesting to relate the findings made in this study with those published in the MutS-DNA crystal studies (Obmolova *et al* 2000; Lamers *et al* 2000). The observed DNA kink in the crystal structure studies seems to be an important hallmark of MutS recognition. It is entirely possible that the kink is protein induced and the intrinsic global flexibility of the heteroduplex might be overridden by MutS induced effects. Moreover, in the co-crystal structures, MutS-bound mismatch is clearly intrahelical showing no signs of "flip-out" and the chemical reactivity results reported here are consistent with this finding. Based on the conclusion emerging from this study, namely the critical combined roles of sequence context and the local chemistry of a mismatch, our future studies are aimed at capturing the dynamic states of MutS-DNA recognitions that accomplish differential affinity to different mismatches. Work is in progress to study fluorescence lifetimes of mismatches *vis-a-vis* MutS binding.

Acknowledgements

We would like to thank Dr Leroy Worth, for providing the MutS clone. We are grateful to Mr Dandekar for help with PhosphorImager analysis and Jatin for photographic prints. We thank Subhojit Sen for helpful discussions.

References

Allawi H T and SantaLucia J Jr 1998 NMR solution structure of a DNA dodecamer containing GT mismatches; *Nucleic Acids Res*. **26** 4925–4934

Arnold F H, Wolk S, Cruz P and Tinoco I 1987 Structure, dynamics, and thermodynamics of mismatched DNA oligonucleotide duplexes d(CCCAGGG)2 and d(CCCTGGG)2; *Biochemistry* **26** 4068–4075

Bhattacharya A and Lilley D 1989a Single Base Mismatches in DNA Long- and short-range structure probed by analysis of axis trajectory and local chemical reactivity; *J. Mol. Biol.* **209** 583–597

Bhattacharya A and Lilley D 1989b The contrasting structures of mismatched DNA sequences containing looped-out bases bulges and multiple mismatches (bubbles); *Nucleic Acids Res.* **17** 6821–6840

Biswas I and Hsieh P 1997 Interaction of MutS protein with the major and minor grooves of a heteroduplex DNA; *J. Biol. Chem.* **272** 13355–13364

Boulard Y, Cognet J A and Fazakerley G V 1997 Solution structure as a function of pH of two central mismatches C·T and C·C in the 29 to 39 K-ras gene sequence by nuclear magnetic resonance and molecular dynamics; *J. Mol. Biol.* **268** 331–347

Brown J, Brown T and Fox K R 2001 Affinity of mismatch–binding protein MutS for heteroduplexes containing different mismatches; *Biochem. J.* **354** 627–633

Fazakerley G V, Quignard E, Woisard A, Guschlbauer W, van der Marel, van Boom J H, Jones M and Radman M 1986 Structures of mismatched base pairs in DNA and their recognition by the *Escherichia coli* mismatch repair system; *EMBO J.* **13** 3697–3703

Gervais V, Cognet J A, Le Bret M, Sowers L C and Fazakerley G V 1995 Solution structure of two mismatches A·A and T·T in the K-ras gene context by nuclear magnetic resonance and molecular dynamics; *Eur. J. Biochem.* **228** 279–290

Hagerman P J 1988 Flexibility of DNA; *Annu. Rev. Biophys. Biophys. Chem.* **17** 265–286

Hunter W N, Brown T, Kneale G, Anand N N, Rabinovich D and Kennard O 1987 The structure of guanosine-thymidine mismatches in B-DNA at 2·5-A resolution; *J. Biol. Chem.* **262** 9962–9970

Jiricny J 1998 Replication errors: cha(lle)nging the genome; *EMBO J.* **17** 6427–6436

Jiricny J, Su S, Wood S and Modrich P 1988 Mismatch-containing oligonucleotide duplexes bound by the *E. coli* mutS-encoded protein; *Nucleic Acids Res.* **16** 7843–7853

Joshi A, Sen S and Rao B J 2000 ATP-hydrolysis-dependent conformational switch modulates the stability of MutS-mismatch complexes; *Nucleic Acids Res.* **28** 853–861

Jones M, Wagner R and Radman M 1987 Repair of a mismatch is influenced by the base composition of the surrounding nucleotide sequence; *Genetics* **115** 605–610

Kahn J D, Yun E and Crothers D M 1994 Detection of localized DNA flexibility; *Nature* (*London*) **368** 163–166

Karthikeyan G, Wagle M and Rao B J 1998 Non-Watson-Crick base pairs modulate homologous alignments in RecA pairing reactions; *FEBS Lett.* **425** 45–51

Kouchakdjian M, Li B F L, Swann P F and Patel D J 1988 Pyrimidine·pyrimidine base pair mismatches in DNA. A nuclear magnetic resonance study of T·T pairing at neutral pH and C·C pairing at acidic pH in dodecanucleotide duplexes; *J. Mol. Biol.* **202** 139–155

Kolodner R 1996 Biochemistry and genetics of eukaryotic mismatch repair; *Genes Dev.* **10** 1433–1442

Kolodner R D and Marsischky G T 1999 Eukaryotic DNA mismatch repair; *Curr. Opin. Genet. Dev.* **9** 89–96

Lamers M H, Perrakis A, Enzlin J H, Winterwerp H H, deWind N and Sixma T K 2000 The crystal structure of DNA mismatch repair protein MutS binding to a G × T mismatch; *Nature* (*London*) **407** 711–717

Lane A, Ebel S and Brown T 1994 Properties of multiple G·A mismatches in stable oligonucleotide duplexes; *Eur. J. Biochem.* **220** 717–727

Lane A N and Peck B 1995 Conformational flexibility in DNA duplexes containing single G·G mismatches; *Eur. J. Biochem.* **230** 1073–1087

Leonard N J, McDonald J J, Henderson R E L and Reichmann A 1971 Reaction of diethyl pyrocarbonate with nucleic acid components. Adenosine; *Biochemistry* **10** 3335–3342

Malkov V A, Biswas I, Camerini-Otero D and Hsieh P 1997 Photocross-linking of the NH_2-terminal region of Taq MutS protein to the major groove of a heteroduplex DNA; *J. Biol. Chem.* **272** 23811–23817

Marra G and Schar P 1999 Recognition of DNA alterations by the mismatch repair system; *Biochem. J.* **338** 1–13

Marathias V M, Jerkovic B and Bolton P H 1999 Damage increases the flexibility of duplex DNA; *Nucleic Acids Res.* 1854–1888

Marathias V M, Jerkovic B, Arthanaari H and Bolton P H 2000 Flexibility and curvature of duplex DNA containing mismatched sites as a function of temperature; *Biochemistry* **39** 153–160

Modrich P and Lahue R 1996 Mismatch repair in replication fidelity, genetic recombination, and cancer biology; *Annu. Rev. Biochem.* **65** 101–133

Mol C D, Parikh S S, Putnam C D, Lo T P and Tainer J A 1999 DNA repair mechanisms for the recognition and removal of damaged DNA bases; *Annu. Rev. Biophys. Biomol. Struct.* **28** 101–128

Obmolova G, Ban C, Hsieh P and Yang W 2000 Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA; *Nature* (*London*) **407** 703–710

Parker B O and Marinus M G 1992 Repair of DNA heteroduplexes containing small heterologous sequences in *Escherichia coli*; *Proc. Natl. Acad. Sci. USA* **89** 1730–1734

Patel D J, Kozlowski S A, Ikuta S and Itakura K 1984a Deoxyguanosine-deoxyadenosine pairing in the d(C-G-A-G-A-AT-T-C-G-C-G duplex: conformation and dynamics at and adjacent to the dG X dA mismatch site; *Biochemistry* **23** 3207–3217

Patel D J, Kozlowski S A, Ikuta S and Itakura K 1984b Deoxyadenosine-deoxycytidine pairing in the d(C-G-C-G-A-A-T-T-C-A-C-G) duplex: conformation and dynamics at and adjacent to the dA X dC mismatch site; *Biochemistry* **23** 3218–3226

Patel D J, Kozlowski S A, Ikuta S and Itakura K 1984c Dynamics of DNA duplexes containing internal G·T G·A A·C and T·C pairs: hydrogen exchange at and adjacent to mismatch sites; *Fed. Proc.* **43** 2663–2670

Plum G E and Breslauer K J 1994 DNA lesions. A thermodynamic perspective; *Ann. NY Acad. Sci.* **726** 45–56

Rajski S R, Jackson B A and Barton J K 2000 DNA repair: models for damage and mismatch recognition; *Mutat. Res.* **447** 49–72

Rayssiguier C, Thaler D S and Radman M 1989 The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch repair mutants; *Nature* (*London*) **342** 396–401

Sixma T K 2001 DNA mismatch repair: MutS structures bound to mismatches; *Curr. Opin. Struct. Biol.* **11** 47–52

Su S and Modrich P 1986 *Escherichia coli* mutS-encoded protein binds to mismatched DNA base pairs; *Proc. Natl. Acad. Sci. USA* **83** 5057–5061

Su S, Lahue R S, Au K G and Modrich P 1988 Mispair specificity of methyl-directed DNA mismatch correction *in vitro*; *J. Biol. Chem.* **263** 6829–6835

Vincze A, Henderson R E L, McDonald J J and Leonard N J 1973 Reaction of diethyl pyrocarbonate with nucleic acid components Bases and nucleosides derived from guanine, cytosine, and uracil; *J. Am. Chem. Soc.* **95** 2677–2682

Werntges H, Steger G, Riesner D and Fritz H J 1986 Mismatches in DNA double strands: thermodynamic parameters and their correlation to repair efficiencies; *Nucleic Acids Res.* **14** 3773–3790

Worth L, Bader T, Yang J and Clark S 1998 Role of MutS ATPase activity in MutS, L-dependent block of *in vitro* strand transfer; *J. Biol. Chem.* **273** 23176–23182

Wu and Crothers D 1984 The locus of sequence-directed and protein-induced DNA bending; *Nature* (*London*) **308** 509–518

MS received 19 September 2001; accepted 21 November 2001

Corresponding editor: DESIRAZU N RAO