Nucleic Acids Research, 1996, Vol. 24, No. 4 721–729

Recognition of DNA insertion/deletion mismatches by an activity in *Saccharomyces cerevisiae*

Juan J. Miret, Breck O. Parker+ and Robert S. Lahue*

Department of Biochemistry and Molecular Biology, University of Massachusetts Medical Center, Worcester, MA 01655, USA

Received September 28, 1995; Revised and Accepted January 3, 1996

ABSTRACT

An activity in nuclear extracts of *S.cerevisiae* binds specifically to heteroduplexes containing four to nine extra bases in one strand. The specificity of this activity (IMR, for insertion mismatch recognition) in band shift assays was confirmed by competition experiments. IMR is biochemically and genetically distinct from the *MSH2* dependent, single base mismatch binding activity. The two activities migrate differently during electrophoresis, they are differentially competable and their spectra of mispair binding are distinct. Furthermore, IMR activity is observed in extracts from an *msh2⁻ msh3⁻ msh4⁻* strain. IMR exhibits specificity for insertion mispairs in two different sequence contexts. Binding is influenced by the structure of the mismatch since an insertion with a hairpin configuration is not recognized by this activity. IMR does not result from single-strand binding because single-stranded probes do not yield IMR complex and single-stranded competitors are unable to displace insertion heteroduplexes from the complex. Similar results with intrinsically bent duplexes make it unlikely that recognition is conferred by a bend alone. Heteroduplexes bound by IMR do not contain any obvious damage. These findings are consistent with the idea that yeast contains a distinct recognition factor, IMR, that is specific for insertion/deletion mismatches.

INTRODUCTION

DNA mismatches may be thought of in two classes. The first class consists of single base mispairs (G/T, for example). In general, the correction of single base mismatches in both prokaryotes and eukaryotes follows a hierarchy of repair specificity where G/T is most efficiently repaired and C/C mispair is poorly repaired (1–3). The remaining mispairs fall between G/T and C/C and are repaired with efficiencies that vary with sequence context and cell species. Several proteins have been implicated in the specific recognition of single base mismatches. The best characterized mismatch recognition protein is *E.coli* MutS which binds mispairs with nanomolar affinities (4,5). Five yeast genes have been identified that bear homology to *mutS* (6–9). These genes, called *MSH* for mutS homolog, encode distinct activities. Strains harboring *msh1* mutations exhibit a deficiency in mitochondrial function (10), *msh2* disruptions yield strong mitotic mutator and post-meiotic segregation (PMS) phenotypes(10,11), *msh3* strains are weak mitotic mutators and give limited PMS (7) and *msh4* and *msh5* disruptions give reduced spore viability and are reduced for meiotic crossing over (8,9). Biochemical analysis of mismatch binding activity in yeast nuclear extracts indicates that the predominant binding activity for single base mispairs is dependent on *MSH2* function (12), consistent with genetic data (10,11) and work with purified Msh2 (13,14).

The second class of mispairs consists of insertion mispairs (extra nucleotides on one strand relative to the other. We refer to this class as insertion mispairs for purposes of clarity because both insertion and deletion mispairs are manifested as extra nucleotides on one strand and hence provide similar targets for repair). These heteroduplex molecules can arise as intermediates during recombination or result from DNA polymerase slippage during replication. In yeast for example, slippage within runs of dinucleotide repeats generates alterations of 2–20 base pairs (bp) (15,16). Deletion events between short repeats separated by 24–147 bp are strongly stimulated by a temperature sensitive DNA polymerase δ mutation, also consistent with a slippage mechanism (17). Several lines of evidence indicate that there may be significant differences in recognition and repair of small versus large insertion mispairs. In bacterial systems, insertion mismatches of up to four nucleotides are corrected *in vivo* whereas 5–514 base mismatches are refractory to repair (18–21). However, large insertions can undergo co-correction if placed near a well-repaired mismatch (21) suggesting that the large insertions escape recognition as opposed to being inherently uncorrectable. This idea is supported by findings that MutS recognizes insertion mispairs up to four bases but not more (20). Tetrad analysis of yeast *msh2* mutants (10,11) suggests that correction of insertion deletions up to four bases requires *MSH2* function. Further support for this idea comes from a study of the destabilization of dinucleotide repeat sequences in yeast (16). Mutations in *MSH2* confer large increases in destabilization due to the inability of these mutants to repair insertion mispairs arising from replicational slippage. Interestingly, destabilization events in the mutant

* To whom correspondence should be addressed

+Present address: Hybridon Incorporated, 1 Innovation Drive, Worcester, MA 01605, USA

strains were limited to additions or deletions of 2–4 bp. As noted (16), this implies that an *MSH2*-dependent activity does not recognize insertion or deletions mismatches that exceed four bases. However a recent report indicates binding of purified yeast Msh2 to oligonucleotides containing certain insertion mispairs up to 14 bases in size (14). Thus there is some uncertainty as what recognizes large insertion mispairs in yeast.

This issue is of importance because eukaryotes are able to correct large insertion mismatches, defined here as DNA containing five or more unpaired bases, as judged by transformation assays (22,23) or by gene conversion experiments. For example, gene conversions in yeast occur with sequences as large as 6.5 kb Ty elements (24,25). Many elements in the 20–1000 bp range undergo gene conversion (25–28). One exception is the 38 bp *ade8-18* deletion which yields looped mispairs that are refractory to correction *in vivo* (1,26). In general there is no bias against the strand containing extra bases since conversion of these large elements often occurs with approximate parity (24,25). Structural properties of larger insertion mismatches affect repair efficiencies. For example, the ability of the extra bases to assume a hairpin configuration may confer resistance to gene conversion. Elegant work by Petes and colleagues has shown that in yeast palindromic insertions capable of forming a hairpin with stem length of at least five base pairs are refractory to repair (27,28).

Recent work in mammalian systems indicates that recognition of single-base and small insertion mispairs is conferred by a heterodimer of hMsh2 and a 160 kDa protein, GTBP (29,30). This complex also restores mismatch correction activity to extracts defective in either Msh2 function or GTBP function (29). Yeast harbors a GTBP homolog, Msh6, that is currently being characterized by Kolodner and colleagues (31). Mutants in *MSH6* have phenotypes that resemble GTBP mutants, suggesting that mismatch recognition and repair activity in yeast may likewise be conferred by a heterodimer. It is not known if Msh2/GTBP is responsible for recognition of large insertion mispairs. Three possibilities exist. First, the Msh2/GTBP heterodimer in eukaryotes provides recognition of large insertion mismatches; second, another MutS homolog provides this function; or third, recognition is conferred by a novel activity. We have investigated insertion mismatch binding in nuclear extracts of *S.cerevisiae*. In this study we present evidence for a previously unknown binding activity, IMR, that specifically recognizes insertion mispairs of 4–9 bases and which exhibits several characteristics that would be expected of a genuine mismatch repair activity.

Figure 1. Oligonucleotide duplexes used as binding substrates. Duplex substrates were prepared by annealing each top strand, 5′ end-labelled with ^{32}P, with its appropriate partner (see Materials and Methods). Two different sequences were utilized to assess context effects. Sequences in the *upper* part of the figure are derived from the *mnt* gene of bacteriophage P22 (32). The heteroduplexes are referred to in the text by the number of extra bases present in the top strand (nine base insertion mismatch, seven base insertion mismatch, etc.) or by the presence of a single base mispair (T/G mispair). The second set of duplexes (*bottom*) corresponds to a portion of the *ADE8* gene of *S.cerevisiae* [nucleotides 119–156, relative to the translation start site (33)]. The *ADE8* modifier is used in the text whenever necessary for clarity. These duplexes are referred to in the text as seven base insertion mismatch, hairpin, T/G, bent duplex, four base insertion mispair and homoduplex, respectively.

MATERIALS AND METHODS

Oligonucleotides and annealing conditions

HPLC-purified oligonucleotides were purchased from Operon Technologies (Alameda, CA). 5′ end-labelled oligonucleotides were prepared using [γ-^{32}P]ATP and polynucleotide kinase. One strand of each duplex (typically the top strand shown in Figure 1) was labelled. Annealing of the 5′ end-labelled oligonucleotides with their complementary strands at a molar ratio of 1:5 was performed in 20 mM Tris–HCl pH 7.6, 10 µM EDTA, 5 mM $MgCl_2$, 100 µM DTT. The mixture was heated to 70°C for 10 min and allowed to cool slowly at room temperature for 30 min. Most experiments utilized 27–36mer oligonucleotides (numbers refer to duplex region) derived from the *mnt* gene (20,32). Additional experiments were performed using 38–39mer oligonucleotides derived from the *ADE8* gene (33). All DNA concentrations are expressed as moles of substrate molecules.

Yeast strains and yeast nuclear extract preparation

Strains used in this study were either DY6 (*MAT*a *ura3-52 leu2 trp1 prb1-1122 pep4-3 prc1-407*; from B. Jones, Carnegie-Mellon University via T. Hsieh, Duke University) or isogenic derivatives (12). Gene disruptions to yield an *msh2::URA3 msh3::TRP1 msh4::LEU2* derivative were performed by single-step protocols (34) using knockout plasmids kindly provided by

R. Kolodner (Dana Farber Cancer Institute), G. Crouse (Emory Univ.) and S. Roeder (Yale Univ.), respectively.

Yeast cell cultures were grown at 30°C in YPD medium to A_{550} = 1.0–2.0, harvested by centrifugation and nuclear extracts were prepared according to the method of Wang *et al.* (35). In some cases, nuclei were extracted with 200 mM NaCl and the resulting proteins were precipitated by addition of ammonium sulfate to a final concentration of 40%. This protocol was found to enrich IMR activity by ~10-fold relative to published methods (35). The final protein concentration of the extracts was 5–25 mg/ml as measured using Lowry assays following precipitation with trichloroacetic acid. Extracts were prepared and tested 2–3 times for each strain. Some variability in insertion binding activity was observed from preparation to preparation so the amount of extract used in the binding assays was adjusted for each experiment to yield similar amounts of insertion complex formation.

Band shift experiments

Binding was performed at 0°C in a total volume of 10 µl containing 20 mM Tris–HCl pH 7.6, 10 µM EDTA, 5 mM $MgCl_2$, 100 µM DTT and 20 µg/ml poly (dI-dC). In the standard assay, cell extracts (final protein concentration 0.5–3.0 mg/ml) were added to 2.5–5 nM radiolabelled oligonucleotide on ice. Extracts from wild-type cells were used unless otherwise stated. When present, unlabelled competitor (200–400 nM, depending on the experiment) was added prior to addition of the extract. Following a 30 min incubation at 0°C, 3.2 µl of 50% (w/w) sucrose was added to the reaction mixture and one-half of the sample was loaded onto a 6% polyacrylamide gel. Electrophoresis was performed at 90 V in 40 mM Tris-acetate, pH 7.5, 2 mM EDTA buffer at 4°C until the bromophenol blue dye front of a control sample had migrated between 7–9 cm (~4–5 h). The gels were subsequently dried and exposed to X-ray film (Kodak X-Omat). In some cases band intensities were quantitated by laser densitometry of the autoradiographs (Biomed Instruments Model SLR-1D/2D).

Analysis of bound and unbound oligonucleotide

200 fmol of the five base insertion mispair in the *mnt* context (see Fig. 1) was incubated with 40 µg of wild-type extract in a reaction volume of 40 µl under standard conditions. Following incubation, the mixture was applied to a preparative band-shift gel and the components separated by electrophoresis. Locations of the IMR-bound and unbound DNA were discovered by autoradiography of the wet gel. Appropriate gel slices were excised and the DNA eluted. Briefly, the gel slices were soaked overnight at room temperature in 20 mM Tris–HCl, pH 7.5, 1 mM EDTA, 150 mM NaCl. Eluted material was applied to an Elutip and eluted in the same buffer containing 1 M NaCl. Following ethanol precipitation, samples were denatured at 95°C for 5 min and applied to an 8% sequencing gel.

RESULTS

Detection of insertion mispair recognition activity in nuclear extracts

Mismatch recognition is an essential step in DNA mismatch correction that occurs when one or more proteins specifically interact with the mismatch and elicit the repair response. Since several lines of evidence indicate that insertion mismatches up to 1000 bases or more are corrected in yeast (25–28), we reasoned that there must be a binding activity that recognizes insertion mismatches. To search for such a binding activity *in vitro* we utilized a band shift assay and oligonucleotides probes with or without a mismatch (Fig. 1). For physiological relevance, the probes were derived from sequences that have been tested *in vivo* and many of which are substrates for mismatch correction. One set of oligonucleotides utilized sequences from the bacteriophage P22 *mnt* gene (32) which was used to study the correction of insertion mismatches in *E.coli* (20,21) as well as mispair binding by the bacterial MutS protein (20). The second set of oligonucleotides was designed within the yeast *ADE8* sequence (33), since Fogel and colleagues (1,26) have proven that deletions and single base mispairs in this region undergo repair. Furthermore, single base mispairs and a one base insertion mismatch placed in either the *mnt* or *ADE8* contexts are recognized by an *MSH2* dependent activity (12).

Radiolabelled duplexes were incubated with yeast nuclear extracts and then subjected to analysis on non-denaturing polyacrylamide gels. It is important to note that the yeast nuclear extracts used in this study were prepared specifically with the intent of revealing all possible mismatch binding activities. This was accomplished by treating purified nuclei with high salt buffer to promote the release and subsequent extraction of as many nuclear proteins as possible. As a result, the crude nuclear extracts produced more than one band when subjected to band shift analysis. We sought to identify insertion mismatch binding activities in these extracts. Specificity of binding was judged by three criteria: requirement for a mismatch, competition for complex formation in the presence of excess mismatch-containing oligonucleotides and reproducibility of binding between extract preparations. Only binding events that satisfied all three requirements were considered.

Incubation of wild-type yeast nuclear extracts with homoduplex DNA (Fig. 2, lane 1) resulted in the formation of several distinct complexes. These represented mismatch-independent binding events and served as controls. Lane 2 shows the result of incubating a one base insertion heteroduplex with nuclear extracts. In addition to the complexes seen before, two new complexes were observed. Competition experiments determined that the upper complex (marked as ‘Single-base complex’) also specifically recognizes the one base insertion mispair; an excess of unlabelled one base insertion heteroduplex displaced the radioactive probe (lane 3) but an excess of non-radioactive five base insertion heteroduplex (lane 4) or homoduplex (lane 5) did not affect complex formation. The characterization of the single-base complex has been established in previous work (12) and corresponds to the *MSH2*-dependent mismatch binding activity. The lower complex (migrating about two-thirds the way down the gel) was not competable and so was not considered further.

The presence of a distinct complex was observed when a probe containing an insertion mismatch of five bases was incubated with nuclear extracts (lane 6). This new complex migrated in a different position from the *MSH2*-dependent complex. To test the specificity of this new complex, several competition experiments were performed. Addition of unlabelled five base heteroduplex reduced complex formation (lane 8) whereas a homoduplex competitor had little effect (lane 9). Note that other complexes further down the gel were competed by homoduplex and hence

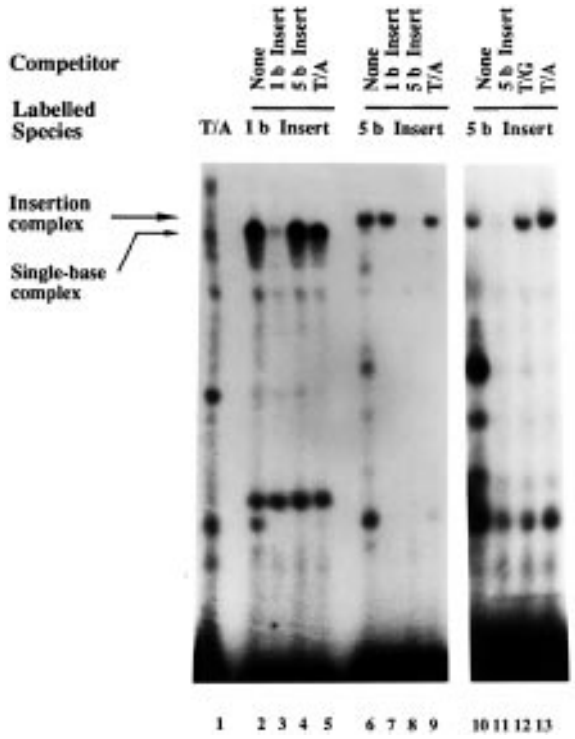

Figure 2. Yeast extracts contain a new binding activity specific for insertion/deletion mismatches. Binding of 50 fmol of radiolabelled duplexes from the *mnt* series (Fig. 1) occurred in a 10 µl reaction containing 200 ng of poly (dI-dC). Nuclear extract containing 5 µg protein (wild-type extract, lanes 1–9) or 30 µg protein (*msh2 msh3 msh4* triple mutant extract, lanes 10–13) was added last. When present, the amount of unlabelled competitor was 4 pmol (lanes 1–9) or 2 pmol (lanes 10–13). Following incubation for 30 min at 0°C, the DNA species were resolved in a 6% polyacrylamide gel under non-denaturing conditions at 4°C. Radiolabelled DNA species were visualized by autoradiography. Unbound DNA migrated to the position at the bottom of the figure. *Arrows* at the side of the figure indicate the positions of species referred to in the text.

do not satisfy this important requirement for mispair specificity. Interestingly, competition with an excess of one base insertion mismatch also failed to prevent specific binding to the five base insertion probe (lane 7). Taken together these results suggest the presence of an activity that recognizes the larger insertion mismatch and is biochemically distinct from the *MSH2*-dependent activity. We have named this new activity IMR, for insertion mismatch recognition.

IMR is independent of MSH2, MSH3 and MSH4

To test the effect of the *mutS* homolog genes *MSH2*, *MSH3* and *MSH4* on mismatch binding, a triple mutant strain with disruptions in all three genes was constructed. Nuclear extracts were prepared from the mutant derivative and mismatch binding was assayed. (Comparison of IMR activity in several extract preparations indicated that apparent differences in the amount of binding activity were due to variations from preparation to preparation and hence these differences result from methodology rather than the genotype of the cell source.) Incubation with a T/G probe (not shown) did not yield any *MSH2* dependent activity in the mutant extract, as expected. In contrast IMR activity was observed with the five base insertion mispair (Fig. 2, lane 10). This activity could be competed by unlabelled five base competitor (lane 11) but not by unlabelled T/G or homoduplex (lanes 12–13). We conclude that IMR does not require function of the *MSH2*, *MSH3* or *MSH4* genes.

IMR activity is largely independent of sequence context

A mismatch recognition activity should be largely independent of DNA sequences flanking the mismatch. To test the influence of flanking DNA sequences, we measured interactions of IMR with seven base insertion mismatches in either the *mnt* or *ADE8* sequence contexts (Fig. 1). The results of this experiment are shown in Figure 3. Lanes 1–7 demonstrate IMR binding to the seven base heteroduplex in the *mnt* context and lanes 8–14 show the result when the mispair occurred within *ADE8*. In the absence of any competitor, IMR activity was observed with both probes (lanes 1 and 8). Complex formation was almost entirely blocked when competitors harboring the seven base heteroduplex in either the *mnt* or *ADE8* sequences were used (lanes 2, 3, 9 and 10). However IMR binding was only modestly reduced by T/G heteroduplex or homoduplex competitors in either sequence context (lanes 4–7 and 11–14). The additional band running above the insertion complex does not exhibit competition patterns that conform to our criteria of specificity and so was not further investigated. The results of Figure 3 support the hypothesis that IMR constitutes a general insertion mismatch binding activity that is largely independent of flanking DNA sequence.

The insertion mismatch complex formation is influenced by structure of the mispair

As an initial assessment of the physiological relevance of IMR activity, we compared its binding properties to two different types of insertion mismatches, one that is repaired *in vivo* and another that is not. Gene conversions of many insertion mutations have been observed (25) including several within *ADE8* (26). However, insertions able to form hairpin structures with stem length of at least five base pairs are poor substrates for repair (27,28). The latter finding is due in part to the fact that, unlike cruciforms arising within palindromic DNA, hairpins within insertion heteroduplexes are not in competition with alternative, highly stable double-strand DNA forms (27,28). The results of incubating nuclear extract with hairpin or seven base insertion heteroduplexes embedded within *ADE8* sequence context are shown in Figure 4. Comparison of lanes 1 and 2 indicates that the hairpin probe (lane 1) gave rise to a complicated pattern of bands that was difficult to interpret. However little material was observed at the IMR position (lane 2), suggesting that the hairpin heteroduplex is not recognized by IMR. To strengthen this finding, competition experiments were performed. IMR binding to the seven base insertion heteroduplex was largely inhibited when unlabelled seven base insertion mispair was the competitor (lane 3). However addition of the hairpin competitor, like the homoduplex competitor, resulted in little reduction of the insertion mispair binding complex (lanes 4–5). The results from Figure 4 support the idea that mispair structure is likely an important feature influencing the recognition and ultimately the repair of such insertion mismatches, consistent with *in vivo* results (27,28).

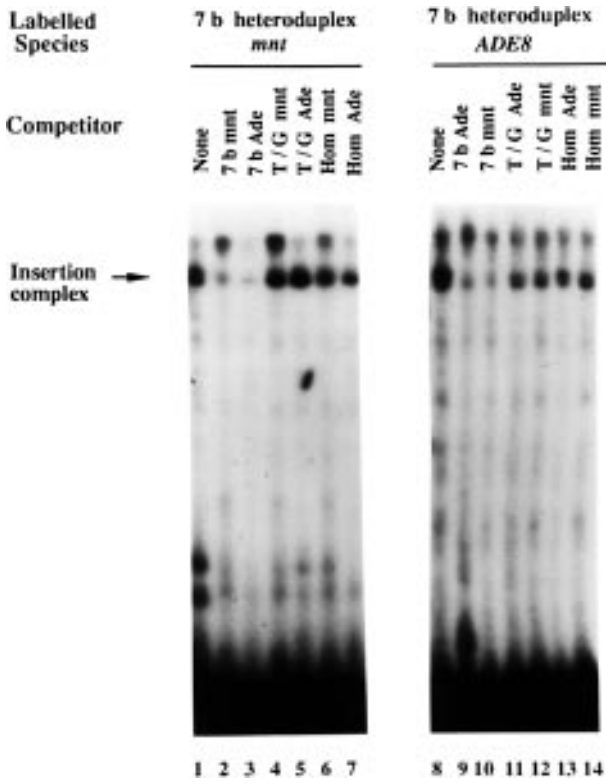

Figure 3. Sequence context effects on mismatch binding activity. Yeast nuclear extract (5 µg) was incubated in a total reaction volume of 10 µl with 25 fmol radiolabelled duplexes derived from either *ADE8* (A) or *mnt* (M) (described in Fig. 1). Unlabelled competitor duplexes (2 pmol), either from *ADE8* (A) or *mnt* (M), were included during the binding reaction as indicated. The figure shows the results of electrophoretic fractionation of the mixtures. The arrows indicate the position of the insertion specific complex described in Figure 2. Densitometric analysis of the insertion complex band intensities indicated that competition for the labelled *mnt* heteroduplex was equally effective for both the unlabelled *mnt* and *ADE8* heteroduplexes (lanes 2 and 3). Modest reductions of approximately equal intensities were detected in lanes 11–14 (T/G and homoduplex competitors from *ADE8* and *mnt*) relative to lane 8 (uncompeted *ADE8* heteroduplex) but these reductions were much less than for the specific competitors (lanes 9–10).

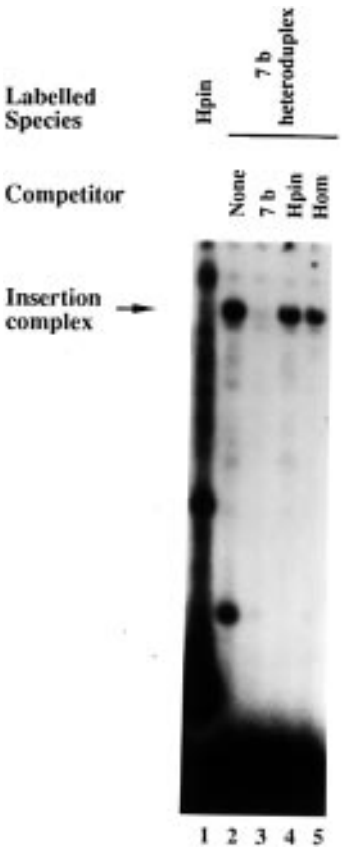

Figure 4. The insertion mismatch complex formation depends on DNA structure. Yeast nuclear extract (10 µg) was incubated in a total reaction volume of 10 µl with 50 fmol radiolabelled duplexes derived from either *ADE8* sequences. Unlabelled competitor duplexes (2 pmol) were included during the binding reaction as indicated. The arrow indicates the position of the insertion specific complex.

IMR is not mediated by single-stranded binding protein or by binding to bent duplex regions

To eliminate the possibility that insertion mismatch binding might be due to single-stranded binding proteins, we radiolabelled both the upper and lower strands that, when annealed to each other, form the nine base insertion complex and repeated the binding assay with these single-stranded probes. Each single-stranded probe gave rise to complexes that migrated at distinctly different positions than the IMR complex (Fig. 5, compare lanes 1 and 2 with lane 3). In addition, single-stranded competitors failed to block complex formation to a nine base heteroduplex probe (lanes 7–8). Competition by a T/G heteroduplex or by homoduplex was also ineffective at blocking IMR complex formation (lanes 5–6). In contrast, a nine base insertion heteroduplex was an effective competitor (lane 4). Note that the IMR complex was the only band exhibiting specific competition in lanes 4–8, again consistent with our criteria of specificity. By inference we ruled out the possibility that all single-stranded competitor in lanes 7–8 was bound by single-stranded binding proteins and was unavailable to challenge formation of the insertion mismatch complex. Since some single-stranded probe remains unbound in lanes 1–2, unlabelled single-stranded competitor must also remain unbound in lanes 7–8 where it is present at 80-fold higher concentration. Thus, free single-stranded competitor is available but it fails to block insertion mismatch complex formation. The results of Figure 5 indicate that single-stranded binding activities are not involved in recognition of the insertion mispairs tested here.

Another possibility for IMR activity is that bends in the DNA duplex are actually being recognized. Insertion mismatches up to four nucleotides contain a bend at the position of the mismatch (36–40). A bent structure might be recognized without regard to the presence of a mispair, thus yielding a result in our assay that was not due to *bona fide* mismatch recognition. To address this possibility, we constructed a duplex containing six consecutive adenine residues within the *ADE8* sequence context (Fig. 1). DNA containing stretches of adenines contain an intrinsic DNA bend (41,42) and hence might act as a target for binding of bend-specific proteins. We assayed the bent molecule for binding and also as a competitor in band shift assays (Fig. 6). Lane 1 shows that the bent DNA probe yielded complexes with nuclear extract that migrated at positions other than the insertion position (lane 2), suggesting that the bent

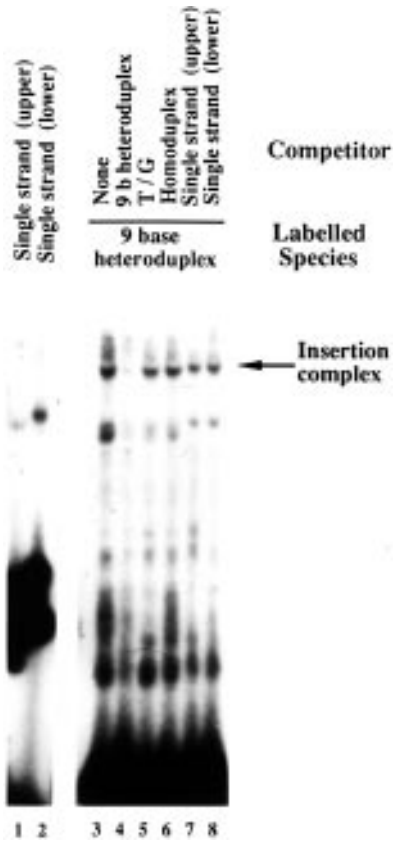

Figure 5. IMR activity is independent of single-stranded binding protein. Binding of nuclear extracts (30 µg) to radiolabelled heteroduplexes was performed under conditions described in Materials and Methods and the products were separated on a polyacrylamide gel. When present, the amount of unlabelled competitor was 4 pmol. The single-stranded probes in lanes 1 and 2 are the oligonucleotides that, when annealed together, form the nine base heteroduplex. 'Upper' and 'lower' refer to their relative positions, as seen in Figure 1. The *arrow* indicates the insertion specific complex. Quantitation of insertion complex band intensities indicated that single-stranded competitors were no more effective at blocking complex formation than were T/G or homoduplex competitors.

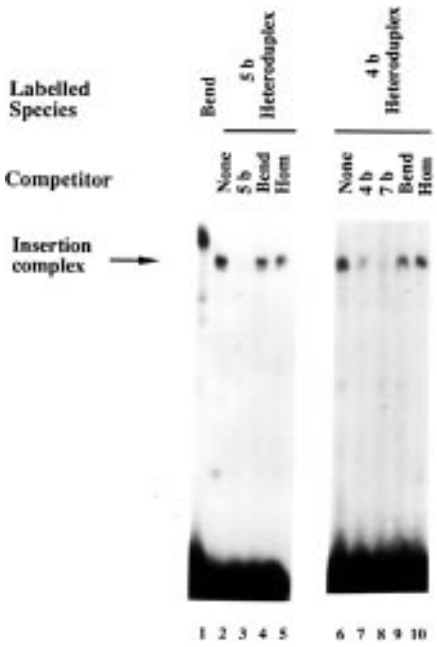

Figure 6. DNA bends are not sufficient to confer recognition by IMR. Binding of nuclear extracts (10 µg) to 50 fmol of radiolabelled duplexes was performed as described in Materials and Methods and the products were separated on a polyacrylamide gel. When present, the amount of unlabelled competitor was 3 pmol. The labelled DNA in lane 1 was the bent duplex (see Fig. 1), in lanes 2–5 it was the five base insertion (*mnt* context) and in lanes 6–10 the probe was the four base insertion mispair (*ADE8* context).

duplex is not being recognized by IMR. In competition experiments the five base insertion probe (*mnt* context) was displaced when excess five base insertion DNA was used (lane 3) but not when the bent DNA or homoduplex was present as the competitor (lanes 4–5). These results were extended to include binding of a four base insertion mispair where the extra nucleotides are all adenine residues (Fig. 1). This mispair contains one strand each from the homoduplex and bent duplex and thus serves as an especially important test of IMR specificity. The four base insertion mispair formed a complex at the IMR position (Fig. 6, lane 6) that was competed by four base insertion (lane 7) or by a seven base insertion in the *ADE8* context (lane 8). In contrast the bent duplex or homoduplex were ineffective at preventing IMR complex formation (lanes 9–10). Essentially identical results (not shown) were obtained in control experiments using extract from an *msh2* mutant strain, excluding the possibility that binding is due to Msh2 protein. Results from genetic analysis of *msh2* mutants (10,11,16) indicate that Msh2 is required for correction of many four base insertion mispairs. The failure to detect *MSH2*-dependent binding to the four base heteroduplex in Figure 6 is unclear. However these findings support the idea that DNA bends are not sufficient to confer recognition by IMR. These results also extend the range of binding of IMR to a four base mispair.

DNA bound by IMR appears unaltered relative to unbound DNA

To assess the possibility that IMR may recognize DNA carrying gross chemical distortions present as a minor subpopulation of our probes, we isolated DNA from the IMR bound and unbound fractions of a band-shift gel like the one shown in Figure 2. The radiolabelled probe was the five base insertion mispair from the *mnt* sequence. The experiment was performed in parallel with either the longer (top) or shorter (bottom) strand of the heteroduplex bearing the radiolabel (Fig. 1). Following isolation the DNA samples were analyzed on a sequencing gel (Fig. 7). Lanes 1–4 provide a sequencing ladder from M13 as a standard. Lanes 5 and 8 contain the unbound and bound fractions, respectively, from the heteroduplex labelled on the shorter strand. There was no significant difference between the two DNA samples. The heterogeneity of the samples in lanes 5 and 8 can be ascribed to exonucleolytic degradation of the two 3′ overhanging nucleotides, with the largest species corresponding to full-length (30mer) oligonucleotide. Similarly, lanes 6 and 7 represent the unbound and bound fractions with the longer strand labelled. Once again, no significant difference is seen and the major component corresponded to the full-length strand (33mer). These results suggest that DNA recognized by IMR does not carry any gross chemical alteration and supports the idea that IMR recognition is conferred by the insertion mispair. The size range of DNA molecules displayed in this gel would have revealed

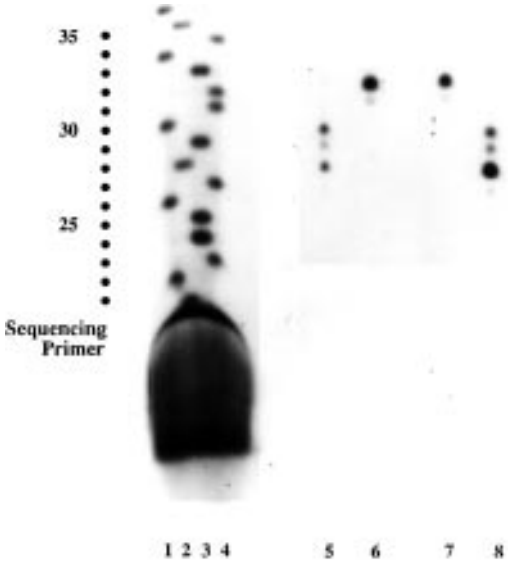

Figure 7. Analysis of bound and unbound insertion mismatch DNA. Five base insertion mispair probe from the *mnt* sequence was prepared with either the longer (top) or shorter (bottom) strand (Fig. 1) bearing a radiolabel. 200 fmol of the five base insertion mispair in the *mnt* context (see Fig. 1) was incubated with 40 µg of wild-type extract in a reaction volume of 40 µl under standard conditions. Following incubation, the mixture was applied to a preparative band-shift gel and the components separated by electrophoresis. Locations of the IMR-bound and unbound DNA were revealed by autoradiography of the wet gel. Appropriate gel slices were excised and the DNA eluted (see Materials and Methods). The figure shows electrophoretic analysis on an 8% sequencing gel. Lanes 1–4 provide a sequencing ladder, loaded in the order G-A-T-C, from an M13 template using a ^{32}P end-labelled '–20' primer (US Biochemicals). Sizes of products (in nucleotides) are shown on the left. The major band at nucleotide position 20 of lanes 1–4 is unextended primer. Lanes 5 and 8, unbound and bound (respectively) fractions of heteroduplex labelled on the shorter strand; lanes 6 and 7, unbound and bound (respectively) fractions of heteroduplex labelled on the longer strand.

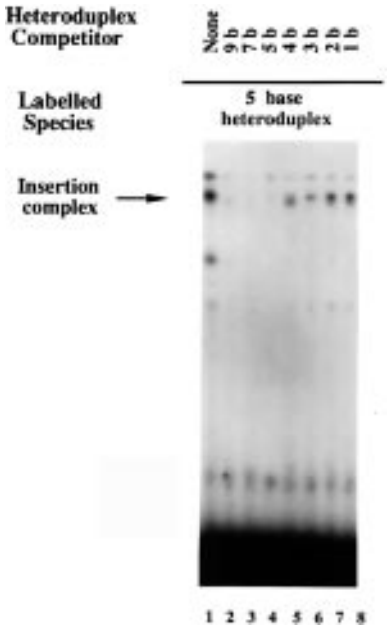

Figure 8. Competition analysis of binding in extracts enriched for IMR. Extracts from an *msh2 msh3* mutant strain were prepared under conditions that led to ~10-fold enrichment of IMR activity (see Materials and Methods). Extract (1.6 µg) was incubated with 25 fmol of labelled five base insertion complex in a volume of 10 µl. When present, 2 pmol of the indicated competitor heteroduplexes were added. All heteroduplexes were from the *mnt* sequences. After incubation, products were separated electrophoretically and detected by autoradiography.

cleavage events at the position of the mispair (13–17 nucleotides from the position of the label). As no cleavage products were observed, we conclude that cleavage is not a required event in IMR binding.

Competition analysis also indicates a range of IMR binding from 4–5 to 9 base insertion mispairs

Based on the results shown above, IMR binding is consistent with the hypothesis of an insertion recognition activity. A systematic analysis of the substrates bound by IMR would provide additional information about its biochemical properties. However attempts to perform this experiment with unfractionated nuclear extracts were difficult to interpret, although the results suggested that IMR may bind insertion mispairs as small as three nucleotides (data not shown). We therefore developed an alternative isolation procedure that enriches our extracts ~10-fold for IMR activity. This enriched extract (from an *msh2 msh3* mutant strain) was subjected to binding analysis using a five base insertion mispair as the labelled probe and unlabelled competitors harboring one to nine base heteroduplexes. Figure 8 shows that complex formation is blocked in the presence of nine, seven or five base competitors (compare lane 1 with lanes 2–4). These results are consistent with observations from experiments shown above for five to nine base heteroduplexes. Competitor heteroduplexes containing one to four bases are largely ineffective at preventing complex formation (lanes 5–8). Heteroduplexes of three or four bases may be at the edge of IMR binding spectrum and of insufficient affinity to compete effectively with the five base probe (Fig. 8), although direct binding to a four base heteroduplex was demonstrated above (Fig. 6).

DISCUSSION

The results of this study show that IMR, a new mismatch recognition activity from yeast, binds specifically to oligonucleotides containing insertion mismatches. This activity has a number of features similar to those of other mismatch binding activities. First, binding is determined by the presence of an insertion mismatch rather than by the DNA sequence context. Second, IMR recognizes a specific group of mismatched oligonucleotides, namely those containing insertion mispairs of four to nine bases. Third, the concentration of probe (nM range) needed for detection of IMR activity is similar to the amount necessary to observe the single base mispair binding activity that requires *MSH2* (12). Fourth, the fact that only a small fraction of the radiolabelled oligonucleotide is bound suggests that, like the *MSH2*-dependent activity (12), IMR may be present in low abundance. Finally, its spectrum of binding is influenced by the structure of the mismatch, as a hairpin structure with a seven base pair stem is not recognized. This binding spectrum correlates well with the available data regarding specificity of repair *in vivo* (27,28). Evidence provided here argues against several modes of recognition that are mismatch independent, including binding to

single-stranded regions, to a bend in the duplex or to DNA carrying grossly damaged regions.

The biochemical and genetic data generated in this study support the idea that IMR is independent of the *mutS* homologs *MSH2*, *MSH3* and *MSH4*. By these criteria, IMR is a previously unidentified activity. We did not test mutants in *MSH1* as this protein has mitochondrial function (10) and purified Msh1 protein exhibits mispair binding spectrum (13) that is not consistent with the insertion binding spectrum described here. We anticipate that *MSH5*, which yields only meiosis-specific mutant phenotypes and falls in the same epistasis group as *MSH4* (9), is unlikely to encode IMR. Attempts to compare directly the spectrum of insertion mispair binding by the *MSH2*-dependent activity and by IMR in our extracts were equivocal due to the complex mixture of proteins present. This characteristic must await comparison of the purified proteins. The range of recognition is of considerable interest with respect to the mechanism (or mechanisms) by which larger insertion mispairs are corrected. Several possibilities for this type of repair are considered below.

The simplest hypothesis is that eukaryotes utilize the same proteins to correct large insertion mispairs that are used in repair of small insertions and single base mispairs. This point hinges on the ability of Msh2/GTBP (Msh2/Msh6 in yeast) to bind insertion mispairs, a property that MutS does not possess (20). A recent report (14) indicates the ability of purified yeast Msh2 to bind insertion mispairs up to 14 nucleotides in length. However it is possible that the insertion mispairs tested, many of which were palindromic, assume some structure that fortuitously provides a single-base mismatch for Msh2 binding. Control binding experiments with bacterial MutS would be most informative on this point. In another study, mammalian cell extracts were shown to support correction of insertion mispairs up to 16 nucleotides (43). Correction of insertion mispairs up to five nucleotides was deficient in extracts from a cell line (44,45) that is mutant in both alleles of *hMSH2*. However correction assays on mispairs of eight or 16 bases were not reported for the *hMSH2* mutant extracts (43). As noted by the authors, there are several explanations for the insertion mismatch repair deficiency in the *hMSH2* mutant extracts and so it remains to be resolved whether this protein participates in correction of large insertion mispairs. Another point that requires resolution with this first hypothesis is that yeast *msh2* mutants yield alterations in dinucleotide repeats of two or four bases (16), suggesting that an *MSH2*-dependent activity fails to recognize insertion or deletion mismatches that exceed four bases. Finally, the identification of a heterodimeric, functional Msh2–GTBP complex in mammalian cells (29,30) and the likelihood of a yeast homolog to GTBP (31) raises the possibility that the predominant mismatch recognition function in yeast is due to a heterodimer, one subunit of which is Msh2. This might explain the difference in heteroduplex specificities between extracts (12) and Msh2 that was overproduced and isolated as a single species (14).

A second possibility is that eukaryotes produce alternate mismatch recognition proteins that are capable of binding to insertion mismatches and then tapping into the same downstream functions that contend with single base and small insertion mismatches (converging pathway hypothesis). As a third possibility, correction of large insertion mismatches may indicate a distinct, previously unknown pathway whose components are not involved with correction of single base mispairs (parallel pathway hypothesis). Our results with IMR are consistent with either of these hypotheses. Several findings suggest that the *mutL* homologs tested to date are not involved in correction of large insertion mispairs. For example, correction in mammalian cell extracts from a cell line defective in *hMLH1* (46,47) was defective in repair of insertions up to four nucleotides (43,48) but exhibited repair activity on insertions of 5–16 nucleotides (43). Also, mutations in *hMLH1* or yeast *MLH1* or *PMS1* yield losses of two bases (49) or gains/losses of two or four bases (16) within dinucleotide repeats, suggesting that at least some of the eukaryotic homologs of *mutL* do not appear to be required for correction of insertion mispairs larger than four nucleotides. However correction of 8 or 12 base palindromic insertions is partially dependent on *PMS1* function (23). Evidence also exists that correction of 27 or 216 nucleotide loops in extracts occurs by a previously unknown pathway (P. Modrich, personal communication; B. Parker and R. Lahue, unpublished results).

One feature of insertion mismatches that influences their correction is secondary structure. For example, different insertion mispairs may assume conformations that are dependent on the exact nature of the inserted bases, leading to recognition by different proteins. If true, this would allow for insertion mispair binding by either IMR or Msh2, depending on the heteroduplex tested and hence could explain the results presented here and by Alani *et al*. (14). For palindromic insertions capable of assuming a hairpin structure with stem length of at least five base pairs, Petes and colleagues have shown that this class of insertion mispairs are not corrected *in vivo*, as judged by gene conversion experiments (27,28). One explanation (28) for the lack of repair of hairpin heteroduplexes is that a protein binds to the hairpin structure and blocks repair. As described in our work, IMR is unable to recognize one of these hairpin structures. Thus the lack of repair of this kind of mismatch could be attributable to a deficiency in recognition rather than a blocking protein. We showed that experiments containing an 80-fold excess of competitor hairpin oligonucleotide was unable to prevent formation of the insertion complex. Unless the blocking protein is present in very large amounts in our extracts, this result suggests that failure of binding to the hairpin by IMR is a more probable explanation for the lack of correction. This conclusion is based on the assumption that IMR is a *bona fide* mismatch correction activity, a supposition that is consistent with the results presented here but which has not yet been directly tested. Unambiguous assays are currently being developed to address this important point.

ACKNOWLEDGEMENTS

We thank Richard Kolodner (Dana Farber Cancer Institute), Gray Crouse (Emory University) and Shirleen Roeder (Yale University) for disruption plasmids for *MSH2*, *MSH3* and *MSH4*, respectively. We thank Martin Marinus (University of Massachusetts Medical Center) for the kind gift of oligonucleotides. This work was supported by National Institutes of Health Grant GM 44824, American Cancer Society Junior Faculty Research Award 327 (to R.S.L.), American Cancer Society Postdoctoral Fellowship 4033 (to B.O.P.) and by funds from the Scientific Council of the University of Massachusetts Medical Center.

REFERENCES

1 Kramer,B., Kramer,W., Williamson,M.S. and Fogel,S. (1989) *Mol. Cell. Biol.*, **9**, 4432–4440.
2 Modrich,P. (1991) *Annu. Rev. Genet.*, **25**, 229–253.
3 Modrich,P. and Lahue,R.S. (1996) *Annu. Rev. Biochem.* **65**, In press.
4 Su,S.-S. and Modrich,P. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5057–5061.
5 Su,S.-S., Lahue,R.S., Au,K.G. and Modrich,P. (1988) *J. Biol. Chem.*, **263**, 6829–6835.
6 Reenan,R.A. and Kolodner,R.D. (1992) *Genetics*, **132**, 963–973.
7 New,L., Liu,K. and Crouse,G.F. (1993) *Mol. Gen. Genet.*, **239**, 97–108.
8 Ross-MacDonald,P. and Roeder,G.S. (1994) *Cell*, **79**, 1069–1080.
9 Hollingsworth,N.M., Ponte,L. and Halsey,C. (1995) *Genes Dev.*, **9**, 1728–1739.
10 Reenan,R.A. and Kolodner,R.D. (1992) *Genetics*, **132**, 975–985.
11 Alani,E., Reenan,R.A. and Kolodner,R.D. (1994) *Genetics*, **137**, 19–39.
12 Miret,J.J., Milla,M.G. and Lahue,R.S. (1993) *J. Biol. Chem.*, **268**, 3507–3513.
13 Chi,N.W. and Kolodner,R.D. (1994) *J. Biol. Chem.*, **269**, 29984–29992.
14 Alani,E., Chi,N.W. and Kolodner,R. (1995) *Genes Dev.*, **9**, 234–247.
15 Henderson,S.T. and Petes,T.D. (1992) *Mol. Cell. Biol.*, **12**, 2749–2757.
16 Strand,M., Prolla,T.A., Liskay,R.M. and Petes,T.D. (1993) *Nature*, **365**, 274–276.
17 Tran,H.T., Degtyareva,N.P., Koloteva,N.N., Sugino,A., Masumoto,H., Gordenin,D.A. and Resnick,M.A. (1995) *Mol. Cell. Biol.*, **15**, 5607–5617.
18 Gasc,A.M., Garcia,P., Baty,D. and Sicard,A.M. (1987) *Mol. Gen. Genet.*, **210**, 369–372.
19 Lefevre,J.-C., Mostachfi,P., Gasc,A.-M., Guillot,E., Pasta,F. and Sicard,M. (1989) *Genetics*, **123**, 455–464.
20 Parker,B.O. and Marinus,M.G. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 1730–1734.
21 Carraway,M. and Marinus,M.G. (1993) *J. Bact.*, **175**, 3972–3980.
22 Bishop,D.K. and Kolodner,R.D. (1986) *Mol. Cell. Biol.*, **6**, 3401–3409.
23 Bishop,D.K., Williamson,M.S., Fogel,S. and Kolodner,R.D. (1987) *Nature*, **328**, 362–364.
24 Vincent,A. and Petes,T.D. (1989) *Genetics*, **122**, 759–772.
25 Petes,T.D., Malone,R.E. and Symington,L.S. (1991) In Broach,J.R., Pringle,J.R., and Jones,E. W., (eds) *The Molecular and Cellular Biology of the Yeast Saccharomyces*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Vol. 1, pp. 407–521.
26 White,J.H., DiMartino,J.F., Anderson,R.W., Lusnak,K., Hilbert,D. and Fogel,S. (1988) *Mol. Cell. Biol.*, **8**, 1253–1258.
27 Nag,D.K., White,M.A. and Petes,T.D. (1989) *Nature*, **340**, 318–320.
28 Nag,D.K. and Petes,T.D. (1991) *Genetics*, **129**, 669–673.
29 Drummond,J.T., Li,G.-M., Longley,M.J. and Modrich,P. (1995) *Science*, **268**, 1909–1912.
30 Palombo,F., Gallinari,P., Iaccarino,I., Lettieri,T., Hughes,M., D'Arrigo,A., Truong,O., Hsuan,J.J. and Jiricny,J. (1995) *Science*, **268**, 1912–1914.
31 Kolodner,R.D. (1995) *Trends Biochem. Sci.*, **20**, 397–401.
32 Rewinski,C. and Marinus,M.G. (1987) *Nucleic Acids Res.*, **15**, 8205–8215.
33 White,J.H., Lusnak,K. and Fogel,S. (1985) *Nature*, **315**, 350–352.
34 Rose,M.D., Winston,F. and Hieter,P. (1988) *Methods in Yeast Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
35 Wang,Z., Wu,X. and Friedberg,E.C. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 4907–4911.
36 Bhattacharyya,A. and Lilley,D.M.J. (1989) *Nucleic Acids Res.*, **17**, 6821–6840.
37 Hsieh,C.-H. and Griffith,J.D. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 4833–4837.
38 Wang,Y.-H. and Griffith,J. (1991) *Biochemistry*, **30**, 1358–1363.
39 Rosen,M.A., Live,D. and Patel,D.J. (1992) *Biochemistry*, **31**, 4004–4014.
40 Rosen,M.A., Shapiro,L. and Patel,D.J. (1992) *Biochemistry*, **31**, 4015–4026.
41 Koo,H.-S., Drak,J., Rice,J.A. and Crothers,D.M. (1990) *Biochemistry*, **29**, 4227–4234.
42 Haran,T.E., Kahn,J.D. and Crothers,D.M. (1994) *J. Mol. Biol.*, **244**, 135–143.
43 Umar,A., Boyer,J.C. and Kunkel,T.A. (1994) *Science*, **266**, 814–816.
44 Fishel,R., Lescoe,M.K., Rao,M.R., Copeland,N.G., Jenkins,N.A., Garber,J., Kane,M. and Kolodner,R. (1993) *Cell*, **75**, 1027–1038.
45 Leach,F.S., Nicolaides,N.C., Papadopoulos,N., Liu,B., Jen,J., Parsons,R., PeltomSki,P., Sistonen,P., Aaltonen,L.A., Nystrsm-Lahti,M. *et al.* (1993) *Cell*, **75**, 1215–1225.
46 Bronner,C.E., Baker,S., Morrison,P.T., Warren,G., Smith,L.G., Lescoe,M.K., Kane,M., Earabino,C., Lipford,J., Lindblom,A. *et al.* (1994) *Nature*, **368**, 258–261.
47 Papadopoulos,N., Nicolaides,N.C., Wei,Y.-F., Ruben,S.M., Carter,K.C., Rosen,C.A., Haseltine,W.A., Fleischmann,R.D., Fraser,C.M., Adams,M.D. *et al.* (1994) *Science*, **263**, 1625–1629.
48 Li,G.-M., Drummond,J., Longley,M.J. and Modrich,P. (1995) Submitted.
49 Parsons,R., Li,G.M., Longley,M.J., Fang,W.H., Papadopoulos,N., Jen,J., de la Chapelle,A., Kinzler,K.W., Vogelstein,B. and Modrich,P. (1993) *Cell*, **75**, 1227–1236.