BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **225,** 289–295 (1996)
ARTICLE NO. 1168

A Carboxy Terminal Domain of the hMSH-2 Gene Product Is Sufficient for Binding Specific Mismatched Oligonucleotides

Adrian Whitehouse,[1] Graham R. Taylor,* Jayne Deeble,* Simon E. V. Phillips,† David M. Meredith, and Alexander F. Markham

*Molecular Medicine Unit, Clinical Sciences Building, *DNA Laboratory, Regional Genetics Service, St. James's University Hospital and †Department of Biochemistry and Molecular Biology, University of Leeds, Leeds, United Kingdom*

Received June 24, 1996

The human MSH-2 gene product is a member of a highly conserved family of proteins which are involved in post-replication mismatch repair. hMSH-2 is homologous to Escherichia coli (E. coli) MutS and Sacchromyces cerevisiae MSH-1 and MSH-2 proteins, which recognise heteroduplex DNA at the sites of all single base mismatches and deletions or insertions up to 4 base pairs. hMSH-2 is one of the hereditary non-polyposis colorectal cancer (HNPCC) tumor supressor genes, and maps to human chromosome 2p16. Alterations in the coding region of the hMSH-2 gene result in a mutator phenotype with marked instability of microsatellite sequences, indicative of a deficiency in DNA repair. It has been shown that purified hMSH-2 binds specifically to nucleotide mismatches in double-stranded DNA. Here we demonstrate that a region of high homology between the members of this class of proteins contains a type A nucleotide binding site consensus sequence which has ATPase activity and is sufficient to bind DNA containing specific mismatched residues.

Hereditary non-polyposis colorectal cancer (HNPCC) is one of the most common human autosomal dominant diseases. Estimations are that up to 1 in 200 of the population may be heterozygotes for mutant alleles and be predisposed to develop this form of cancer (1). HNPCC kindreds are characterized by a high incidence of mutation in tumour microsatellite sequences when these are compared with the normal repeats occurring in non-malignant tissues from the same individuals (2). Also, cell lines derived from HNPCC tumours are genetically unstable (3). These observations suggested that a possible causative factor was a failure of DNA repair. It has now been shown that cancer predisposition in these patients is attributable to defects in any one of four human genes involved in mismatch repair, hMSH-2, hMLH-1, hPMS-1 or hPMS-2 (4-9).

The best studied mismatch repair pathway is the methyl-directed long patch repair pathway in E. coli (reviewed in 10), which is involved in increasing the fidelity of replication by specific repair of DNA polymerase incorporation errors. Initiation of heteroduplex repair is dependent on the product of the MutS gene, which binds to base mispairs and loops of up to four unpaired nucleotides. After this, the MutL gene product binds to the DNA-MutS complex, initiating excision of a tract of single-stranded DNA that contains the mismatched residues. The repair process is completed by resynthesis of the excised DNA strand and ligation of the remaining nick.

Mutations of MutS and MutL, or their homologues in yeast, lead to increases in the rate of spontaneous mutation resulting in a mutator phenotype (11-12). hMSH-2 and the recently discovered G/T binding protein (GTBP), which are believed to form a heterodimer (hMutSα) are homologs of MutS (13-14), whereas hMLH-1, hPMS-1 and hPMS-2 are analogous to MutL. Mutations in any of these genes are believed to inactivate mismatch repair in man and

[1] Corresponding author. Fax: 044 113 2444475. E-mail: whitehouse@stjames.leeds.ac.uk.

0006-291X/96 $18.00

yield a mutator phenotype which can destroy the normal functioning of critical genes and lead to tumour formation (reviewed in 15).

hMSH-2 has been shown to bind to DNA containing nucleotide mismatches in vitro (16-17). hMSH-2 and its homologues are very highly conserved over their carboxy terminal domains. This region of the protein demonstrates homology to the type A consensus sequence found in many proteins that bind and hydrolyse nucleotides (18). It has been shown that MutS displays a weak ATPase activity in the presence and absence of DNA and genetic alteration of this ATP binding site results in a protein which is defective in mismatch repair (19). In this paper, we report the expression and purification of a 241 amino acid fragment from the carboxy terminus of hMSH-2, and show that this domain displays ATPase activity and highly specific mismatch binding activity *in vitro*.

METHODS

Construction of a hMSH-2 C-terminal domain expression vector. A DNA fragment containing the cDNA fragment required from hMSH-2 was generated by polymerase chain reaction (PCR) using 10ng of plasmid pBShMSH-2 DNA (a generous gift of Prof. B. Vogelstein), and 250ng of the oligonucleotides dCCG AAG CTT AGG CAT GCT TGT GTT GAA GTT CAA GAT and dGCG GGA TCC TCT TTC CAG ATA GCA CTT CTT TGC TGC. These oligonucleotides incorporated BamHI and HindIII restriction sites respectively, for convenient cloning of the PCR product. They generate a PCR fragment which encodes amino acid 637 to amino acid 877 in the published sequence (4,20). The reaction was performed with 4 units of Taq polymerase (Promega) in the buffer recommended by the supplier. After 30 cycles (1 min, 92°C, 1 min 60°C, 1 min 72°C), the DNA produced was phenol/chloroform extracted, ethanol precipitated, digested with BamHI and HindIII and cloned into the corresponding sites of pFlag.CTC (IBI) to derive phMSH-2.Flag. The integrity of the insert was checked by DNA sequencing (data not shown).

Expression of hMSH-2 C-terminal domain as a bacterial fusion protein. The PCR product encoding the amino acid 637-877 hMSH-2 domain in the Flag bacterial expression vector (IBI), was used to transform E. coli strain DH5α. A fresh overnight culture of transformed E. coli was diluted 1 in 20 with LB medium containing ampicillin (100 μg/ml). After growth at 37°C for 2 hours, the culture was induced with IPTG (1 mM) and grown at 37°C for a further 5 hours. The cells were harvested by centrifugation at 3200g for 10 minutes and resuspended in 0.1 volume lysis buffer (100 mM Tris, pH 8.0, 1 mM EDTA) and incubated on ice with 3 mg/ml of lysozyme for 30 minutes. The cells were then sonicated and lysed by the addition of Tween 20 lysis buffer (100 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.3 mg/ml phenylmethylsulphonyl fluoride, 0.8 μg/ml pepstatin, 1 mM DTT, 1% Tween 20). Cellular debris was pelleted by centrifugation at 4,000g.

Detection of fusion protein by Western blot analysis. Protein extracts were mixed with 2× reducing sample buffer (50mM Tris-HCl, pH 6.8, 4% sodium dodecyl sulphate (SDS), 5 mM EDTA, 10% β-mercapthoethanol, 1 mM DTT and 0.01% bromophenol blue). After boiling for 3 minutes, samples were fractionated on a 12% SDS polyacrylamide gel. After electrophoresis the gel was soaked for 10 min in transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol (v:v), and 0.1% SDS), and the proteins were transferred to nitrocellulose membranes by electroblotting for 3 hours at 250 mA. After transfer, the membranes were soaked in PBS and incubated for 2 hr in blocking buffer (PBS containing 5% nonfat dry milk). Membranes were incubated with a 1/100 dilution of the M2 monoclonal antibody (IgG_1, IBI), washed with PBS and incubated for 1 hr at 37°C with a 1/1000 dilution of rabbit anti-mouse immunoglobulin conjugated with horseradish peroxidase in blocking buffer. After five washes with PBS the nitrocellulose membranes were developed in PBS containing 0.02% 1-chloro-4-naphthol and 0.006% hydrogen peroxide.

ATPase assay. The assay was performed at 37°C in 20 mM Tris-HCl pH 7.6, 0.5 mM $CaCl_2$, 5 mM $MgCl_2$, 1 mM DTT, 100 μg/ml BSA, 0.1 mM EDTA and 150 ng of hMSH-2 domain. Assays were performed using 2, 2.5, 3.3, 5 and 10 μM ATP. Hydrolysis of [α-^{32}P]ATP by the carboxy terminal domain of hMSH-2 was assayed by thin layer chromatography. The radioactive counts for ATP and its hydrolysis products were quantified using a scintiallation counter (Packard).

Functional binding assay. Mismatch binding was detected by a nitrocellulose binding assay of labelled oligonucleotides followed by autoradiography. Oligonucleotides (dCGG ATC CGG AXG TCA TGG AAT TCC and dGGA ATT CCA TXA CAT CCG GAT CCG) were synthesised and annealed to produce either a perfect matched double-stranded molecule or a single mismatch (position shown in bold type). Oligonucleotides were mixed to a final concentration of 100 pmole/μl each in 100 μl STM (100 mM NaCl, 10 mM Tris-HCl, pH 7.0, 10 mM $MgCl_2$, 5 mM DTT) heated to 95°C and cooled to 25°C over 2 hours. The annealed products were then stored in 50% glycerol at −20°C until required. End-labelling of double-stranded DNA (100 pmole) in STM buffer was by polynucleotide kinase. After incubation at 20°C for 10 minutes the unincorporated label was removed using a Sephadex NAP 5 column. The labelled DNA was diluted to 0.2 pmole/μl. The binding assay used 1 pmole of DNA with 150 ng hMSH-2 domain

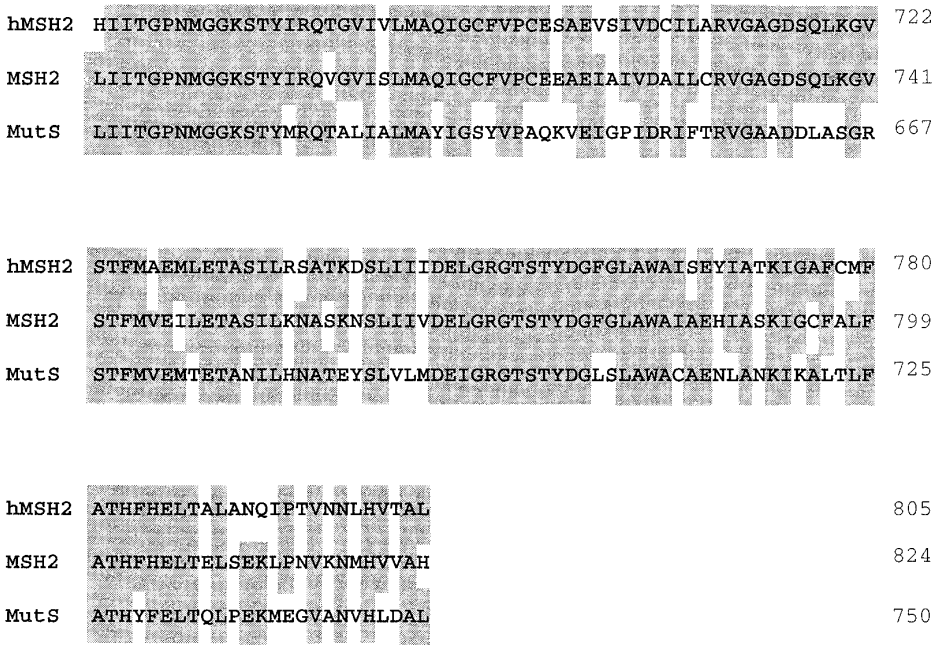

FIG. 1. Alignment of amino acid sequences of the conserved COOH terminal region of hMSH-2, MSH2 and MutS.

in a total volume of 10 μl. After 1 hour on ice the mixture was slowly filtered over pure prewetted nitrocellulose (Millipore, 0.45 μm) and washed in STM buffer. The filter was then allowed to air dry and bound DNA was detected by autoradiography. Bound material was quantified using a scintillation counter (Packard).

RESULTS

PCR Amplification and Cloning

Protein sequence alignments of hMSH-2 and its homologues, MutS, MSH-2 revealed a highly conserved region at the COOH terminus (Fig. 1). This region contains a type A nucleotide binding site consensus sequence. A 720 bp fragment was amplified using PCR, incorporating BamHI and HindIII restriction sites for convenient cloning. This fragment of the hMSH-2 cDNA sequence and encodes amino acid residues 637 to 877. The PCR product was ligated to pFlag.CTC, in phase with respect to the ATG translational start codon immediately upstream of the multiple cloning site (MCS) and also in frame with the C-terminal coding sequence immediately downstream of the MCS to ensure proper fusion to the C-terminal Flag peptide (Asp Tyr Lys Asp Asp Asp Asp Lys), (Fig. 2).

Expression of the hMSH-2 C-Terminal Domain

The hMSH-2 domain was thus cloned into the bacterial expression vector Flag (IBI). Expression of the hMSH-2-Flag fusion protein resulted in a 30 kDa species detected by Western blot analysis on SDS-PAGE (Fig. 3). The anti-Flag M2 monoclonal (IgG1) mouse antibody (IBI) was used to specifically bind to the eight amino acid Flag peptide, which identified the 249 amino acid recombinant protein comprising the hMSH-2 domain (containing a type A nucleotide binding site consensus sequence) coupled to the Flag peptide at its carboxy terminus.

ATPase Analysis of Bacterial Fusion Protein

The Walkers A-type nucleotide binding motif conserved in MutS proteins has been shown to have ATPase activity (19). In order to determine whether the carboxy terminal domain of

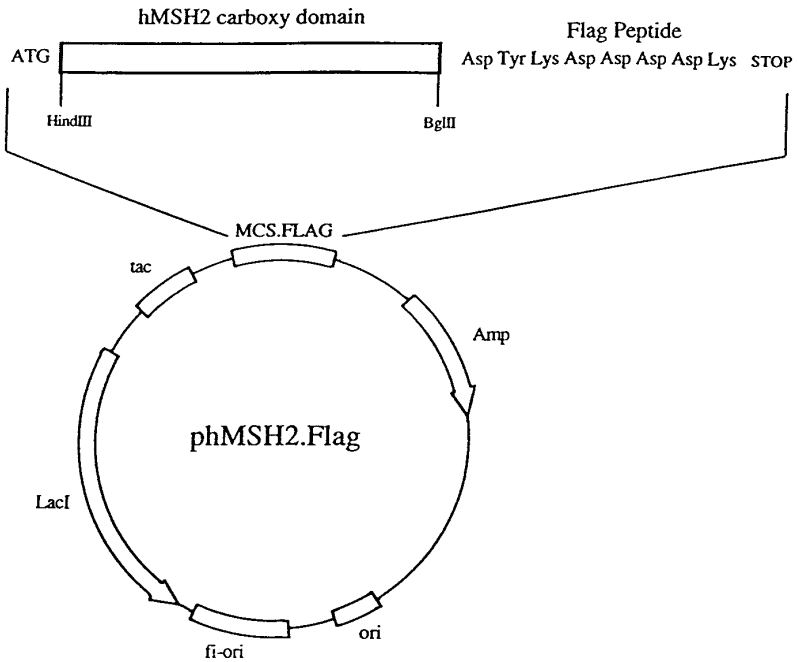

FIG. 2. A DNA fragment containing the carboxy terminal domain of hMSH-2 was generated using PCR. This fragment contained amino acids 637-877 of the published sequence. The domain was ligated to pFlag.CTC to derive phMSH-2.Flag.

hMSH-2 hydrolyses ATP to ADP and Pi, [α^{32}P]ATP was incubated with the fusion protein and separated using TLC. To determine K_m and k_{cat} values of the hMSH-2 domain, ATPase activity was measured in the presence of various concentrations of ATP (Fig. 4). At 37°C the K_m and k_{cat} were calculated to be 6.6 μM and 0.5 s^{-1}, respectively. In a control experiment, nonenzymatic hydrolysis of ATP in the absence of the expressed domain was less than 5%.

Functional Analysis of the Bacterial Fusion Protein

A mismatch binding assay was developed to measure the hMSH-2 C-terminal domain's activity. Mismatch binding was detected by nitrocellulose binding of labelled oligonucleotides containing a mismatch at position 11 within the context of a double-stranded 24-mer oligonucleotide pair. The binding of the hMSH-2 domain to a range of mismatch containing oligonucleotides is shown in Fig. 5. Radiolabelled oligonucleotides containing a perfect match or a single mismatch were incubated with protein extracts containing hMSH-2.Flag or Flag alone. The binding of proteins to mismatched or matched oligonucleotides was quantified using a scintillation counter (Fig. 6). The hMSH-2 domain selectively bound to the oligonucleotides containing all possible mismatches apart from C/C and A/A mismatches, but not the perfectly matched pair. The Flag control did not bind to any of the labelled oligonucleotide, showing the hMSH-2 domain alone is sufficient to bind oligonucleotides containing mismatches.

DISCUSSION

Based on the high degree of sequence conservation between hMSH-2 and its homologues around the domain of the type A nucleotide binding motif, we hypothesized that this region would be sufficient to bind mismatched oligonucleotides. Herein we have expressed this domain as a bacterial fusion protein and shown that a range of mismatch-containing oligonucleotides

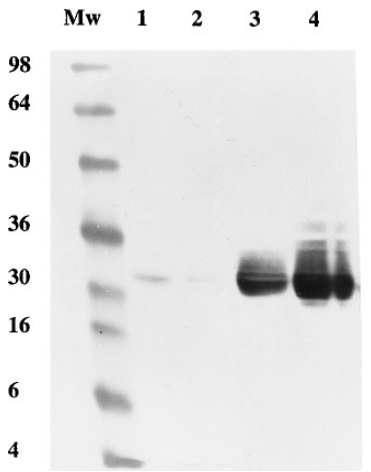

FIG. 3. Analysis of hMSH-2.Flag fusion protein. Transformed *E. coli* were grown at 37°C for 2 hours, the cultures were grown for a further 5 hours (Lane 1), or induced with IPTG (1 mM) and grown for 0 hours (Lane 2), 2 hours (Lane 3) and 5 hours (Lane 4). Extracts were resolved by SDS PAGE, transferred to nitrocellulose, and incubated with M2 monoclonal antibody (IgG_1) and immune complexes were detected by using rabbit anti mouse immunoglobulin conjugated with horseradish peroxidase. The nitrocellulose membranes were developed in PBS containing 0.02% 1-chloro-4-naphthol and 0.006% hydrogen peroxide.

are selectively bound. Also this fusion protein retains ATPase activity which has been shown to be important in mismatch repair (19).

A domain of hMSH-2 was chosen as earlier attempts to express full length protein either in baculovirus or prokaryotic expression systems were unsuccessful (data not shown). This is probably a consequence of the overexpression of full length hMSH-2 being deleterious to growth. However, the 241 amino acid C-terminal domain has now been shown to be expressed at relatively high levels in bacteria. The binding of mismatched oligonucleotides and ATPase activity from crude protein extracts from induced cultures demonstrates that there is specific production of a protein with the predicted characteristics of the domain. Furthermore, as the

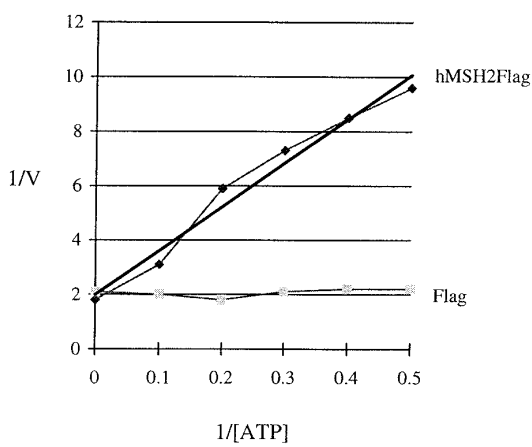

FIG. 4. ATPase analysis of the bacterial fusion. Hydrolysis of various substrate concentrations of [α-^{32}P]ATP by the carboxy terminal domain of hMSH-2 was assayed by thin layer chromatography, and quantified using a scintillation counter.

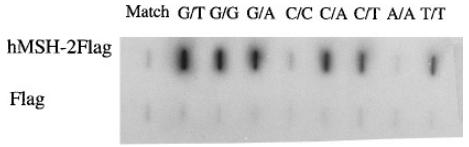

FIG. 5. Functional analysis of the bacterial fusion protein. Oligonucleotides containing either a perfect match or a range of single mismatches were radiolabelled using polynucleotide kinase. One pmole of labelled DNA was incubated for 1 hour with protein extracts of hMSH-2.Flag or Flag. After the incubation period the mixtures were slowly filtered over prewet nitrocellulose, washed and bound DNA was detected using autoradiography.

activities cannot be detected in uninduced cells it is likely that the mismatch binding and ATPase activities are due to the construct alone, rather than a complex formed between the induced protein and endogenous protein.

To date, this protein has proved refractory to purification using anti-Flag antibody beads. The protein only remains soluble in the presence of greater than 1 mM DTT, which is incompatible with immunoaffinity chromatography. In an attempt to overcome this problem we have expressed this domain using the pET expression system. Although high levels of protein were achieved, we were unable to obtain any soluble protein using reagents compatible with either immunoaffinity chromatography or metal affinity chromatography. This domain like the equivalent regions in all other homologues, is hydrophobic. Attempts to overcome the insolubility of the expressed domain through construction of a thioredoxin-hMSH-2 fusion protein led to the failure to increase solubility (data not shown). As expression levels of the hMSH-2Flag fusion are, at present, less than 0.5% of soluble protein, large scale culture and extensive purification will be required to produce sufficient material for further studies.

This domain of hMSH-2 may be of considerable value in the detection of mismatches in DNA generally. New methods are now being investigated based on the use of proof-reading enzymes or repair enzymes. Binding of MutS to mismatches has been shown to be sensitive and specific (21). However, MutS which has been expressed in bacterial systems and purified is rather unstable. We believe that the hMSH-2 domain described and expressed in this paper may be used in similar assays and that it will be more stable.

It has recently been reported that DNA mismatch recognition and binding in human cells is achieved by two distinct proteins, hMSH-2 and GTBP. These two proteins are believed to form a heteroduplex, hMutSα, which is thought to be necessary for correction of base-base mismatches and small insertion/deletions. However, protein sequence alignments show that both proteins are members of the MutS family, which may have been duplicated from a primitive MutS gene. Both proteins are highly homologous in their carboxy terminal domains.

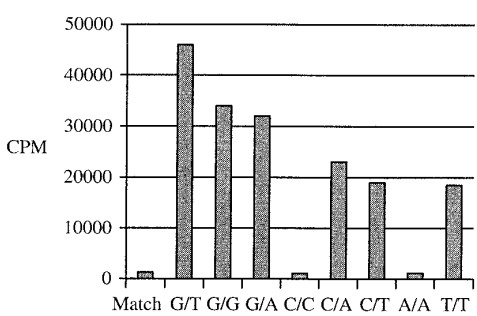

FIG. 6. Quantification of binding assay. Radioactive spots on the nitrocellulose filter were excised and quantified using a scintillation counter. The results are shown as counts per minute.

Therefore, it will be of interest to express the homologous domain of GTBP to establish whether it alone also has mismatch binding properties in the absence of hMSH-2 and to examine the two recombinant domains together to determine whether they function co-operatively.

ACKNOWLEDGMENTS

We thank the Yorkshire Cancer Research Campaign and West Riding Medical Research Trust for financial support for this project. Research in the authors' laboratory is also supported by the Medical Research Council, North Yorkshire Regional Health Authority, and the Wellcome Trust. S.E.V.P. is an International Research Scholar of the Howard Hughes Medical Institute. We are particularly grateful to Pat Clissold for advice and helpful discussions and to Bert Vogelstein for providing the clone pBShMSH-2.

REFERENCES

1. Lynch, H. T., Smyrk, T. C., Watson, P., Lanspa, S. J., Lynch, J. F., Cavalieri, R. J., and Bolland, C. R. (1993) *Gastroentrology* **104,** 1535–1549.
2. Loeb, L. A. (1994) *Cancer Res.* **54,** 5059–5063.
3. Parsons, R., Li, G. M., Longleym, M. J., Fang, W., Papadopoulos, M., Jen, J., de la Chapelle, A., Kinzler, K. W., Vogelstein, B., and Modrich, P. (1993) *Cell* **75,** 1227–1235.
4. Leach, F. S., Nicolaides, N. C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., Peltomaki, P., Sistonen, P., Aaltonen, L. A., Nystrom-Lahti, M., Guan, X.-Y., Zhang, L., Meltzer, P. S., Yu, J.-W., Kao, F.-T., Chen, D. J., Cerosaletti, K. M., Fournier, R. E. K., Todd, S., Lewis, T., Leach, R. J., Naylor, S. L., Weissenbach, J., Mecklin, J.-P., Jarvinen, H., Petersen, G. M., Hamilton, S. R., Green, J., Jass, J., Watson, P., Lynch, H. T., Trent, J. M., de la Chapelle, A., Kinzler, K. W., and Vogelstein, B. (1993) *Cell* **75,** 1215–1225.
5. Bronner, C. E., Baker, S. M., Morrison, P. T., Warren, G., Smith, L. G., Lescoe, M. K., Kane, M., Earabino, C., Lipford, J., Lindbolm, A., Tannergard, P., Bollag, R. J., Godwin, A. R., Ward, D. C., Norden-skjold, M., Fishel, R., Kolodner, R., and Liskay, R. M. (1994) *Nature* **368,** 258–261.
6. Kolodner, R. D., Hall, N. R., Lipford, J., Kane, M. F., Rao, M. R. S., Morrison, P., Wirth, L., Finan, P. J., Burn, J., Chapman, P., Earabino, C., Merchant, E., and Bishop, D. T. (1994) *Genomics* **24,** 516–526.
7. Kolodner, R. D., Hall, N. R., Lipford, J., Kane, M. F., Morrison, P., Wirth, L., Finan, P. J., Burn, J., Chapman, P., Earabino, C., Merchant, E., and Bishop, D. T. (1994) *Cancer Res.* **55,** 242–248.
8. Liu, B., Nicolaides, N. C., Markowitz, S. S., Willson, J. K. V., Parsons, R. E., Jen, J., Papadopoulos, N., Peltomaki, P., de la Chapelle, A., Hamilton, S. R., Kinzler, K. W., and Vogelstein, B. (1995) *Nature Genetics* **9,** 48–55.
9. 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., Venter, J. C., Hamilton, S. R., Petersen, G. M., Watson, P., Lynch, H. T., Peltomaki, P., Mecklin, J.-P., de la Chapelle, A., Kinzler, K. W., and Vogelstein, B. (1994) *Science* **263,** 1625–1629.
10. Modrich, P. (1991) *Annu. Rev. Genet.* **25,** 229–253.
11. Cox, E. C. (1976) *Annu. Rev. Genet.* **10,** 135–156.
12. Strand, M., Prolla, T. A., Liskay, R. M., and Petes, T. D. (1993) *Nature* **365,** 274–276.
13. Drummond, J. T., Li, G.-M., Longley, M. J., and Modrich, P. (1995) *Science* **268,** 1909–1912.
14. Palambo, F., Gallinari, P., Iaccarino, I., Lettieri, T., Hughes, M., D'Arrigo, A., Truong, O. J., Hsuan, J., and Jiricny, J. (1995) *Science* **268,** 1912–1914.
15. Modrich, P. (1994) *Science* **266,** 1959–1960.
16. Fishel, R., Ewel, A., Lee, S., Lescoe, M. K. (1994) *Cancer Res.* **54,** 5539–5542.
17. Fishel, R., Ewel, A., Lee, S., Lescoe, M. K., and Griffith, J. (1994) *Science* **266,** 1403–1405.
18. Walker, J. W., Saraste, M., Runswick, M. J., and Gay, N. J. (1982) *EMBO J.* **1,** 945–951.
19. Haber, L. T., Pang, P. P., Sobell, D. I., Mankovich, J. A., and Walker, G. C. (1988) *J. Bacteriol.* **170,** 197–202.
20. Fishel, F., Lescoe, M. K., Rao, M. R. S., Copeland, N. G., Jenkins, N. A., Garber, J., Kane, M., and Kolodner, R. (1993) *Cell* **75,** 1027–1038.
21. Ellis, L. A., Taylor, G. R., Banks, R., and Baumberg, S. (1994) *Nucleic Acids Research* **22,** 2710–2711.