(R_S = 30.5 Å). Purified hMutSα eluted at a value for R_S of 67 Å. Band sedimentation through 10 to 30% sucrose gradients in buffer A was performed at 4°C at 68,000*g* for 24 hours in a Beckman SW 50.1 rotor. Sedimentation markers run in parallel tubes were catalase [sedimentation coefficient ($s_{20,w}$) = 11.3*S*], γ-globulin ($s_{20,w}$ = 7.1*S*), and ovalbumin ($s_{20,w}$ = 3.55*S*). Sedimentation relative to these markers yielded a value of $s_{20,w}$ for hMutSα of 9.0*S*.

15. L. M. Siegel and K. J. Monty, *Biochim. Biophys. Acta* **112**, 346 (1966).
16. E. J. Cohn and J. T. Edsall, in *Proteins, Amino Acids, and Peptides as Ions and Dipolar Ions*, E. J. Cohn and J. T. Edsall, Eds. (Hafner, New York, 1943), pp. 370–381.
17. Peptide sequence analysis was performed as described (*9*), after resolution of subunits (about 100 pmol) by electrophoresis in the presence of SDS. Attempts to obtain the NH_2-terminal sequence of either subunit were unsuccessful; however, sequence information was obtained for seven internal tryptic peptides of p160. Three of these were unambiguous, and the mass of each peptide was consistent with the mass predicted from the sequence. These were PT48 (GGHFYSALPEILR), PT56 (VHVQFFDDSPTR), and PT64 (TLLEEEYFR). Two others gave secondary sequences in early cycles, and only the unambiguous sequence information is given here: PT36 (DAAWSEAGPGPR) and PT40 (SVLEGDPSENYR). Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
18. S.-S. Su, R. S. Lahue, K. G. Au, P. Modrich, *J. Biol. Chem.* **263**, 6829 (1988).
19. N.-W. Chi and R. D. Kolodner, *ibid.* **269**, 29984 (1994).
20. E. Alani, N.-W. Chi, R. Kolodner, *Genes Dev.* **9**, 234 (1995).
21. M. Grilley, K. M. Welsh, S.-S. Su, P. Modrich, *J. Biol. Chem.* **264**, 1000 (1989).
22. A. Umar, J. C. Boyer, T. A. Kunkel, *Science* **266**, 814 (1994).
23. G.-m. Li, M. J. Longley, J. T. Drummond, unpublished results.
24. W.-h. Fang and P. Modrich, *J. Biol. Chem.* **268**, 11838 (1993).
25. V. S. Goldmacher, R. A. Cuzick, W. G. Thilly, *ibid.* **261**, 12462 (1986); A. Kat *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6424 (1993).
26. J. Holmes, S. Clark, P. Modrich, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5837 (1990); D. C. Thomas, J. D. Roberts, T. A. Kunkel, *J. Biol. Chem.* **266**, 3744 (1991).
27. L. T. da Costa *et al.*, *Nature Genet.* **9**, 10 (1995).
28. F. Palombo *et al.*, *Science*, **268**, 1912 (1995).
29. N. Papadopoulos *et al.*, *ibid.*, p. 1915.
30. Cell lines HeLa S_3, H6, S0, TK6, and MT1 were grown as described (*2, 25, 26*). The HCT-15 cell line obtained from the American Type Culture Collection was cultured in roller bottles (850 cm^2) in RPMI 1640 medium supplemented with 20% fetal bovine serum (Hyclone, Logan, UT). LoVo cells were cultured in a similar manner in McCoy's 5A medium containing glutamine (4 mM final concentration), 10% fetal bovine serum, penicillin (100 U/ml), and streptomycin (0.1 mg/ml).
31. Phage f1 heteroduplexes representing eight base-base mispairs and two-, three-, or four-nucleotide loopout, slipped-strand insertion-deletion mismatches were prepared as described (*2, 18*). Bacteriophage f1MR21 and f1MR22 were prepared by insertion of synthetic oligonucleotide duplexes (5′-AGCTGCAGCCAAAAAGGCTTGG-3′ annealed with 3′-CGTCGGTTTTTCCGAACCGATC-5′ or 5′-AGCTGCAGCCAAAAAAGGCTTGG-3′ annealed with 3′-CGTCGGTTTTTCCGAACCGATC-5′, respectively), into f1MR1 replicative form DNA (*18*) that had been cleaved with Hind III and Xba I. Resulting isolates were used to prepare heteroduplexes containing ⟨A⟩ and ⟨T⟩ insertion-deletion mismatches within overlapping restriction sites for Bgl I and Xcm I. Mismatch repair assays (*2, 26*) were performed with 50 μg of nuclear extract and 24 fmol (0.1 μg) of heteroduplex DNA in 10 μl at 37°C for 15 min. All results are averages of at least two determinations. For band shift assays, oligonucleotide duplexes containing a G-T mispair, a ⟨T⟩ insertion mismatch, a ⟨TTT⟩ insertion mismatch, or a control A-T base pair were formed by annealing radiolabeled 5′-^{32}P-GCTAGCAAGCTTTCGATTCTAGAAATTCGGC-3′ with unlabeled 5′-GCCGAATTTCTAGAATCGAGAGCTTGCTAGC-3′, 5′-GCCGAATTTCTAGAATCGAAGCTTGCTAGC- 3′, 5′-GCCGAATTTCTAGAATCGGCTTGCTAGC-3′, or with 5′-GCCGAATTTCTAGAATCGAAAGCTTGCTAGC-3′, respectively.
32. J.T.D. isolated and characterized hMutSα and demonstrated that HCT-15 and MT1 cells are deficient in this activity. G.-M.L. identified the selective repair defects in HCT-15 and MT1 cell lines and demonstrated that hMutSα binds to mismatched base pairs.

We thank F. Leach, D. Hill, and B. Vogelstein for monoclonal antibody EH12; E. Korytynski and S. Larson for culturing the cell lines; B. Vogelstein and J. Jiricny for sharing unpublished results; and the Harvard Microchemistry Facility for peptide analysis. Supported by grant GM45190 from the National Institute of General Medical Sciences, translational funds from the Duke Comprehensive Cancer Center, and by an American Cancer Society postdoctoral fellowship PF-3940 (M.J.L.).

6 January 1995; accepted 4 May 1995

GTBP, a 160-Kilodalton Protein Essential for Mismatch-Binding Activity in Human Cells

Fabio Palombo, Paola Gallinari, Ingram Iaccarino, Teresa Lettieri, Melya Hughes, Antonello D'Arrigo, Oanh Truong, J. Justin Hsuan, Josef Jiricny*

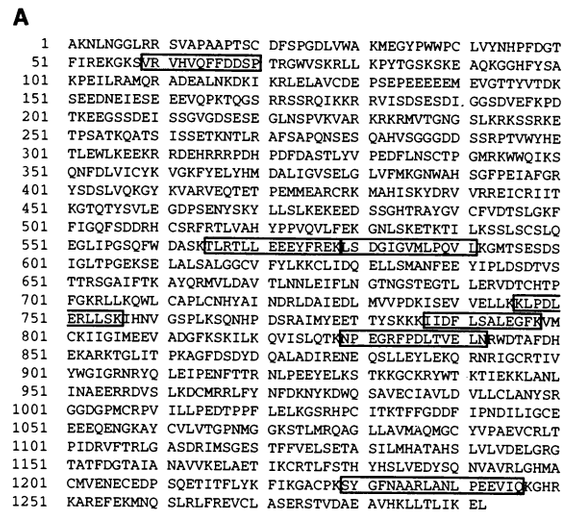
DNA mismatch recognition and binding in human cells has been thought to be mediated by the hMSH2 protein. Here it is shown that the mismatch-binding factor consists of two distinct proteins, the 100-kilodalton hMSH2 and a 160-kilodalton polypeptide, GTBP (for G/T binding protein). Sequence analysis identified GTBP as a new member of the MutS homolog family. Both proteins are required for mismatch-specific binding, a result consistent with the finding that tumor-derived cell lines devoid of either protein are also devoid of mismatch-binding activity.

Cells from human hereditary nonpolyposis colorectal cancers (HNPCC) exhibit a mutator phenotype with a marked instability of microsatellite sequences, a phenotype that has been traced to defects in DNA mismatch repair [reviewed in (*1*)]. The serendipitous discovery of open reading frames (ORFs) encoding murine and human polypeptide homologs of the *Escherichia coli* mismatch-binding protein MutS (*2*) paved the way for the identification of an ever-growing family of MutS homolog (MSH) genes (*3*). Proteins encoded by three members of this family (*Saccharomyces cerevisiae* MutS homologs MSH1 and MSH2, and the human homolog hMSH2) have been shown to bind to DNA mismatches in vitro (*4*). The link between the biological function of hMSH2 and the phenotype of the HNPCC tumors was forged when the *hMSH2* gene was shown to segregate with a known HNPCC locus on chromosome 2p (*5*), and the *hMSH2*-deficient colorectal tumor cell line LoVo was shown to be deficient in mismatch repair (*6*) as well as in mismatch-binding activity (*7*) and to exhibit a marked instability of microsatellite sequences (*8*).

We identified a mismatch-binding factor in HeLa cells (*9*) that binds preferentially to heteroduplexes containing G/T mispairs and one- and two-nucleotide loops (*7*). Purification of this DNA binding activity by G/T mismatch affinity chromatography (*10*) yielded a mixture consisting of a 100-kD protein that we later showed to be hMSH2 (*11*) and a second protein with an apparent molecular weight of 160 kD. As it was this latter polypeptide that became covalently bound to the G/T substrate in cross-linking experiments, we named it GTBP (*10*). To identify GTBP and elucidate its role in mismatch binding, we subjected a purified sample to proteolysis and microsequence analysis (*12*). Seven peptide sequences were obtained: VRVHVQFFDD, KLPDLERLXSK, LSRGIGVMLPQVL, TLDTLLEEEYFREK, SYGFNAARLANLPEEVIQ, NPEGRFPDLTVELN, and IIDFLSALEGFK (*13*). Following the strategy described in (*14*), we identified a unique DNA sequence encoding the central eight amino acids of the peptide SYGFNAARLANLPEEVIQ (*13*). We then prepared a double-stranded DNA probe based on this sequence and used it to screen a HeLa complementary DNA (cDNA) library in phage lambda (*15*). The longest identified clone, C1, contained an insert of 3992 nucleotides with an ORF encoding a polypeptide of 1292 amino acids and a predicted molecular mass of 142 kD (Fig. 1A). As all seven peptides obtained from the microsequence analysis of the 160-kD protein were located within this ORF, we concluded

F. Palombo, P. Gallinari, I. Iaccarino, T. Lettieri, M. Hughes, A. D'Arrigo, J. Jiricny, Istituto di Ricerche di Biologia Molecolare P. Angeletti, Via Pontina km 30.600, 00040-Pomezia, Italy.
O. Truong and J. J. Hsuan, Ludwig Institute for Cancer Research, University College London Medical School, 91 Riding House Street, London W1P 8BT, UK.
*To whom correspondence should be addressed.

Fig. 1. Predicted amino acid sequence of GTBP. (**A**) Sequence of the protein encoded by clone C1. The numbers on the left indicate amino acid position. (*13*). The peptides identified by microsequencing are boxed. The first amino acid residue of construct FLY5 (*15*) is Asn[116]. The cDNA sequence has been deposited in GenEmbl Bank (accession no. U28948). (**B**) Alignment of the amino acid sequences of the conserved COOH-terminal regions of the mismatch-binding proteins GTBP (*Homo sapiens*), hMSH2 (*H. sapiens*), MSH2 (*S. cerevisiae*), and MutS (*E. coli*). Conserved amino acid residues are boxed. The ATP binding site consensus sequences (*3*) are indicated by roman numerals I to IV. Sequences reported in the alignment correspond to entries MSH2YEAST (MSH2) and MUTSECOLI (MutS) in the SwissProt databank or to the coding region of GenBank entry HSU04045 (hMSH2). The alignment was carried out with the use of the GCG Pileup program. The figure was generated with the use of the Prettyplot option.



that clone C1 encodes GTBP. Northern (RNA) blot analysis revealed that the mRNA encoding the 160-kD protein is approximately 4.2 kb in length (*16*), which indicates that our cDNA was incomplete.

Computer search of protein databanks revealed that the GTBP amino acid sequence is highly homologous to that of the MSH family of proteins, particularly at the COOH-terminus (Fig. 1B). GTBP can thus be considered a new member of the MSH family.

Selective antisera (*17*) were then used to investigate whether hMSH2 and GTBP bind as a complex to DNA mismatches. Preincubation of HeLa cell nuclear extracts with rabbit antiserum to either protein, but not with preimmune sera, before addition of the G/T heteroduplex probe inhibited the formation of the protein-DNA complex (Fig. 2A), suggesting that the mismatch-binding factor consists of both proteins. This finding implied that extracts from cells lacking either protein should be devoid of mismatch-binding activity. Thus, we could confirm an earlier finding (*7*) that LoVo cells, which contain a homozygous deletion in both *hMSH2* alleles, contain no mismatch-binding activity (Fig. 2B). This activity was absent also from another colorectal cancer cell line, DLD1 (Fig. 2B), in which neither *hMSH2* allele appears to be mutated (*18*). Two lines of evidence indicate that the DLD1 cells are devoid of functional GTBP: (i) No full-length GTBP was detected in DLD1 extracts by immunoblot analysis (Fig. 2C). (ii) Sequence analysis of GTBP cDNA from DLD1 cells showed that both alleles are inactivated by frameshift mutations (*19*). Together, these data strongly suggest that the mismatch-binding factor in human cells is composed of two proteins, hMSH2 and GTBP.

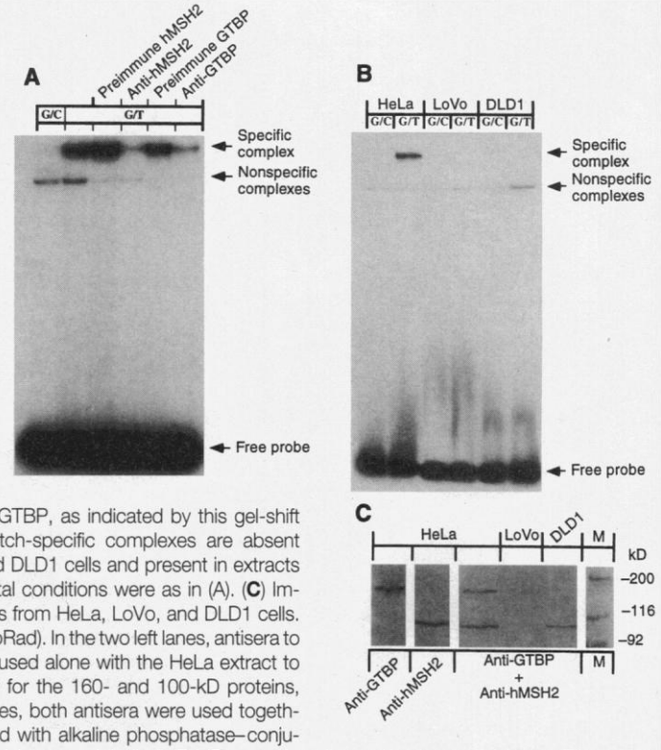


Fig. 2. (**A**) Effect of antisera to hMSH2 (anti-hMSH2) and to GTBP (anti-GTBP) on the formation of the specific mismatch-binding complex. This gel-shift analysis was carried out as described (*9*), except that nuclear extracts were used (*23*). The antisera were added to the reaction mixtures 20 min before the radioactively labeled probe. The figure is an autoradiogram of a native 6% polyacrylamide gel run in tris acetate EDTA (TAE) buffer. (**B**) Mismatch-binding activity is absent from cell extracts lacking hMSH2 or GTBP, as indicated by this gel-shift assay showing that mismatch-specific complexes are absent from extracts from LoVo and DLD1 cells and present in extracts from HeLa cells. Experimental conditions were as in (A). (**C**) Immunoblot analysis of extracts from HeLa, LoVo, and DLD1 cells. M, molecular size marker (BioRad). In the two left lanes, antisera to GTBP and to hMSH2 were used alone with the HeLa extract to demonstrate their selectivity for the 160- and 100-kD proteins, respectively. In the other lanes, both antisera were used together.The bands were visualized with alkaline phosphatase–conjugated goat antibodies to rabbit immunoglobulin G (Fc) (Promega) as directed by the manufacturer. The amount of hMSH2 and GTBP in DLD1 and LoVo extracts, respectively, was considerably lower than that in the HeLa cell extracts, possibly because the two proteins are unstable when not in a complex (*10*).

We then expressed the two polypeptides in a cell-free translation system and assayed their ability to bind mismatch-containing substrates. We used a hMSH2 cDNA clone and the GTBP clones C1 and FLY5 (*15*) as starting materials for the in vitro transcription and translation reactions (Fig. 3A), which were carried out as in (*20*). Gel-shift assays (Fig. 3B) revealed that mismatch-specific binding activity was dependent on the expression of both hMSH2 and GTBP. Thus, although neither protein alone was active in this assay, mixing of the hMSH2 and GTBP translation products increased the mismatch-specific binding activity. We confirmed this result with GTBP cDNA clone FLY5, which encodes a truncated GTBP protein (*15*). As shown in Fig. 3B, mixing of the hMSH2 and FLY5 translation products with the G/T probe produced a new band with faster electrophoretic mobility than that of the endogenous complex, as would be expected of a smaller species.

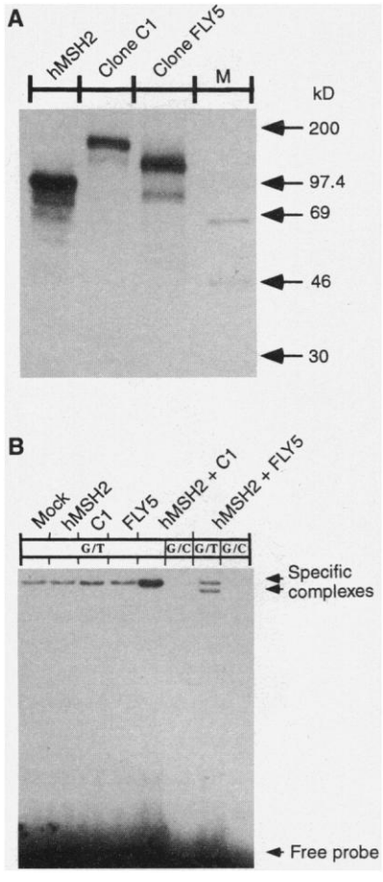


Fig. 3. Reconstitution of mismatch-binding activity with in vitro translated GTBP and hMSH2. (**A**) Translation of hMSH2, GTBP (clone C1), and FLY5 mRNAs in a reticulocyte lysate system (Promega) produced polypeptides of 113, 144, and 122 kD, respectively (*24*). The figure is an autoradiogram of a denaturing 7.5% SDS-polyacrylamide gel. M, molecular size marker (Amersham). (**B**) Gel-shift analysis, performed as in Fig. 2, showing the binding of the in vitro–translated proteins to the G/T heteroduplex. The figure is an autoradiogram of a native 6% polyacrylamide gel run in TAE buffer.

What is the biological role of the hMSH2-GTBP complex? In an independent study (*21*), Modrich and collaborators isolated a mismatch-binding factor from HeLa cells, which they call hMutSα, that is identical to the hMSH2-GTBP complex described here. Their experiments indicate that this complex is necessary for the correction of base-base mispairs and one- or two-nucleotide loops, but that it may not be absolutely required in the correction of larger insertions and deletions such as those associated with microsatellite instability (*21*). Their findings are substantiated by data from Vogelstein and collaborators (*19*), who have shown that the genomic instability in tumor cell lines devoid of GTBP is manifested primarily as small rather than large changes in CA repeats, changes that are associated with mutations in the four known HNPCC loci *hMSH2*, *hMLH1*, *hPMS1*, and *hPMS2*. A number of tumors display mutator phenotypes with a similarly low degree of microsatellite instability (*22*). Future studies will reveal whether these tumors are defective in GTBP.

REFERENCES AND NOTES

1. P. Modrich, *Science* **266**, 1959 (1994)
2. J. P. Linton *et al.*, *Mol. Cell. Biol.* **9**, 3058 (1989); H. Fujii and T. Shimada, *J. Biol. Chem.* **264**, 10057 (1989).
3. L. New, K. Liu, G. F. Crouse, *Mol. Gen. Genet.* **239**, 97 (1993).
4. N.-W. Chi and R. D. Kolodner, *J. Biol. Chem.* **269**, 29984 (1994); T. A. Prolla, Q. Pang, E. Alani, R. D. Kolodner, R. M. Liskay, *Science* **265**, 1091 (1994); E. Alani, N.-W. Chi, R. Kolodner, *Genes Dev.* **9**, 234 (1995); R. Fishel, A. Ewel, M. K. Lescoe, *Cancer Res.* **54**, 5539 (1994).
5. R. Fishel *et al.*, *Cell* **75**, 1027 (1993); F. S. Leach *et al.*, *ibid.*, p. 1215.
6. A. Umar *et al.*, *J. Biol. Chem.* **269**, 14367 (1994).
7. G. Aquilina *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 8905 (1994).
8. D. Shibata *et al.*, *Nat. Genet.* **6**, 273 (1994).
9. J. Jiricny *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8860 (1988).
10. M. Hughes and J. Jiricny, *J. Biol. Chem.* **267**, 23876 (1992).
11. F. Palombo *et al.*, *Nature* **367**, 417 (1994).
12. GTBP was purified as in (*10*). The fractions containing the G/T-specific mismatch-binding activity were loaded onto a preparative SDS-polyacrylamide gel and the 100- and 160-kD bands were excised after electrophoresis and staining with Coomassie blue. The proteins were digested in the gel with trypsin (100-kD protein; Promega) or with *Achromombacter lyticus* endopeptidase lys-C (160-kD protein; Wako Chemicals, Neuss, Germany). The proteolytic peptides were recovered by sequential extractions and separated by tandem high-performance liquid chromatography on a Hewlett-Packard 1090M system fitted with a diode array detector. Anion-exchange and octadecyl reversed-phase columns were connected in series [as in H. Kawasaki and K. Suzuki, *Anal. Biochem.* **186**, 264 (1990)]. Fractions were collected and applied directly to an Applied Biosystems 477A pulsed-liquid automated sequencer [modified as in N. F. Totty, M. D. Waterfield, J. J. Hsuan, *Protein Sci.* **1**, 1215 (1992)].
13. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln, R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
14. J. Lingner, J. Kellermann, W. Keller, *Nature* **354**, 496 (1991).
15. Two degenerate primers corresponding to the NH_2- and COOH-terminal amino acid sequences of SYGFNAARLANLPEEVIQ (*13*) (5′-GCGAATTC-TAYGGNTTYAAYGC-3′ and 5′-GCGGATCCTAYTGDATNACYTC-3′, where N = any nucleotide; Y = C or T; and D = A, G, or T) were used for polymerase chain reaction (PCR) amplification of polyadenylated HeLa mRNA as in (*14*), except that the $MgCl_2$ concentration was 5 mM. The expected 67–base pair (bp) fragment was eluted from an acrylamide gel, cloned into pGEM-3Zf$^+$, and sequenced. Two clones contained the correct sequence encoding the central eight amino acid residues (NAARLANL) (*13*) of the starting target peptide. On the basis of this sequence, we designed a double-stranded probe: 5′-TATGGGTTTAATGCAGCAAGGCTTGCTAATCT-3′ and 5′-TTCCGAACGATTAGAGGGGTCTCCTTCAATAGGTTCT-3′. The probe was labeled by filling-in with Klenow polymerase and deoxythymidine triphosphate (dTTP), dGTP, dCTP, and [α-^{32}P]dATP, and was used to screen an oligo(dT)-primed HeLa S3 cDNA library in phage lambda (Uni-ZAP XR, Stratagene). Two positive clones were selected for further analysis. Clone C1 contained an insert of 3992 bp with an ORF encoding amino acids 1 to 1292, and clone FLY 5 contained an ORF encoding amino acids 116 to 1292.
16. F. Palombo and J. Jiricny, data not shown.
17. PCR fragments corresponding to amino acid residues 27 to 158 of hMSH2 and 750 to 928 of GTBP were subcloned into the *E. coli* expression vector pGEX-3X (Pharmacia/LKB), and the recombinant proteins, in the form of fusion polypeptides with glutathione-S-transferase, were isolated as recommended by the manufacturer, except that the final concentration of isopropyl-β-D-thiogalactopyranoside was 0.25 mM, and the induced cultures were harvested after 6 hours at 20°C. The fusion proteins were used for immunization of New Zealand White S.P.F. female rabbits (Charles River) according to standard protocols.
18. L. T. Costa *et al.*, *Nat. Genet.* **9**, 10 (1995); B. Vogelstein, personal communication.
19. N. Papadopoulous *et al.*, *Science* **268**, 1915 (1998).
20. S. M. Powell *et al.*, *N. Engl. J. Med.* **329**, 1982 (1993)
21. J. T. Drummond *et al.*, *Science* **268**, 1909 (1995).
22. L. A. Aaltonen *et al.*, *ibid.* **260**, 812 (1993); S. N. Thibodeau, G. Bren, D. Schaid, *ibid.*, p. 816; Y. Ionov *et al.*, *Nature* **363**, 558 (1993); R. Wooster *et al.*, *Nat. Genet.* **6**, 152 (1994); A. Merlo *et al.*, *Cancer Res.* **54**, 2098 (1994).
23. J. D. Dignam *et al.*, *Methods Enzymol.* **101**, 582 (1983).
24. The C1 and FLY5 ORFs were introduced into pCite-2b (Novagen). The hMSH2 ORF was inserted into pCite-1 (Novagen). In vitro transcription and translation reactions were carried out as described in (*20*) and included a control reaction without added DNA. ^{35}S-labeled translation products were analyzed on an SDS-polyacrylamide gel that was treated with Amplify (Amersham), dried, and autoradiographed. Gel shift was carried out as in (*9*). Five-microliter samples of single in vitro translation reactions were tested. In the premixing experiments, 2.5 ml of each of the two translation reactions were mixed and incubated for 15 min at room temperature before the addition of the probe. Adenosine monophosphate (5 mM) was included in all the DNA binding reactions to overcome the effect of ATP in the reticulocyte lysates [it prevents the formation of mismatch-specific protein-DNA complexes (*10*)].

25. We thank B. Vogelstein for the hMSH2 cDNA clone and for many helpful discussions; P. Modrich for sharing unpublished data; T. Kunkel and M. Bignami for the cell lines; C. Nardi and D. Schmid for help in the initial stages of this study; P. Neuner for oligonucleotides; P. Costa for the antisera; M. Emili for photography; P. Neddermann and A. Lahm for helpful discussions; and R. Laufer and R. Cortese for critical reading of the manuscript.

14 April 1995; accepted 25 May 1995