indicates that local site-specific effects are the same in both configurations. However, these local effects can be sex-specific, because the order changes to 3D-11A-1B-18A in males (Fig. 1). Both observations are consistent with the possibility that the local effects are related to the transcriptional activity of nearby genes.

We interpret the results shown in Figs. 1 and 3 to indicate that the search for homology in *Drosophila* is more effective when the matching sequences are physically connected, even if the distance between them is more than 15 million base pairs. We favor this view because the homology search is the only step in the recombinational repair process with an apparent genome-wide interaction. If repair happens after the S phase of the cell cycle, it is possible that the preferred template is actually on the sister chromatid and is not physically connected to the break. However, the process would still require some long-range mechanism to distinguish that chromatid from either of the two sister chromatids of the homologous chromosome.

Much of the recent work on homologous recombination has used systems in which one or both of the interacting sequences reside on a multicopy extrachromosomal plasmid (*2*, *4*, *5*, *7*, *15*) or in which both sequences are closely linked (*2*, *3*, *6*, *7*). These strategies eliminate the need for a genome-wide homology search and thus would not reveal the long-range cis effect seen here. Other experiments on recombination in yeast have indicated that there is a cis effect over distances of 20 to 70 kb, but it diminishes rapidly over longer ranges (*16*–*18*). It is possible that the very long range effects we observe occur only in species with more complex genomes than *Saccharomyces cerevisiae*. In the present data, even the nearest site, position 3D, is still several hundred kilobases away from the *white* locus. It may be significant that the 3D site also had the highest reversion rates (Figs. 1 and 3).

It is unclear what biological advantage might exist for a homology search mechanism that scans the genome preferentially in cis. One possibility is that the homology search occurs as a succession of increasingly specific steps. Rao and Radding (*19*) have shown that the RecA-catalyzed homology search recognizes both complementary and identical sequences and even sequences in which the DNA backbone is reversed. They suggest that this recognition represents a preliminary scanning to be followed by a more specific comparison. It is possible that the long-range cis effect we observe reflects an even cruder step that occurs early in the refinement process.

REFERENCES AND NOTES

1. N. A. Nassif and W. R. Engels, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1262 (1993).
2. A. S. Waldman and R. M. Liskay, *ibid.* **84**, 5340 (1987); C. Deng and M. R. Capecchi, *Mol. Cell. Biol.* **12**, 3365 (1992).
3. J. Rubnitz and S. Subramani, *Mol. Cell. Biol.* **4**, 2253 (1984); R. M. Liskay, A. Letsou, J. L. Stachelek, *Genetics* **115**, 161 (1987).
4. K. R. Thomas, C. Deng, M. R. Capecchi, *Mol. Cell. Biol.* **12**, 2919 (1992).
5. P. Shen and H. V. Huang, *Genetics* **112**, 441 (1986).
6. N. Sugawara and J. E. Haber, *Mol. Cell. Biol.* **12**, 563 (1992); B. Y. Ahn, K. J. Dornfeld, T. J. Fagrelius, D. M. Livingston, *ibid.* **8**, 2442 (1988).
7. C. Mézard, D. Pompon, A. Nicolas, *Cell* **70**, 659 (1992).
8. W. R. Engels, D. M. Johnson-Schlitz, W. B. Eggleston, J. Sved, *ibid.* **62**, 515 (1990).
9. G. B. Gloor, N. A. Nassif, D. M. Johnson-Schlitz, C. R. Preston, W. R. Engels, *Science* **253**, 1110 (1991).
10. D. M. Johnson-Schlitz and W. R. Engels, *Mol. Cell. Biol.* **13**, 7006 (1993).
11. We determined the conversion rate for each insertion site by crossing males bearing the w^{hd} allele [$w^{hd80k17}$ in previous work (*20*)], one copy of P[*walter*], and the Δ2-3 transposase source (*14*) with C(1)DX,*y f* females (*21*) and scoring the progeny as described (*9*). The w^{hd} allele has a 629-bp P element inserted into exon 6 of the gene (*20*). Its phenotype is bleach-white. Precise excision of the P element results in full reversion of w^{hd} (*8*). There is no measurable reversion of w^{hd} in the absence of P transposase (*8*). The phenotype of P[*walter*] in a *white* mutant background is intermediate between bleach-white and that of the wild type and is easily distinguished from either. For each site, a sample of the revertants was tested by PCR [polymerase chain reaction (*22*)] and restriction mapping as described (*9*) to change reversion rates to conversion rates. Previous results indicate that background reversion of w^{hd} in the absence of a template is essentially absent in females and very rare (0.1%) in males (*8*).
12. B. S. Baker and J. M. Belote, *Annu. Rev. Genet.* **17**, 345 (1983).
13. B. J. Thomas and R. Rothstein, *Cell* **56**, 619 (1989).
14. H. M. Robertson *et al.*, *Genetics* **118**, 461 (1988).
15. V. M. Watt, C. J. Ingles, M. S. Urdea, W. J. Rutter, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4768 (1985); B. S. Singer, L. Gold, P. Gauss, D. H. Doherty, *Cell* **31**, 25 (1982); S. Jeong-Yu and D. Carroll, *Mol. Cell. Biol.* **12**, 5426 (1992).
16. M. Lichten and J. E. Haber, *Genetics* **123**, 261 (1989).
17. M. Kupiec and T. D. Petes, *Mol. Cell. Biol.* **8**, 2942 (1988).
18. G. S. Roeder, M. Smith, E. J. Lambie, *ibid.* **4**, 703 (1984).
19. B. J. Rao and C. M. Radding, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6646 (1993).
20. G. M. Rubin, M. G. Kidwell, P. M. Bingham, *Cell* **29**, 987 (1982).
21. D. L. Lindsley and G. G. Zimm, *The Genome of* Drosophila melanogaster (Academic Press, San Diego, 1992).
22. H. A. Erlich, *PCR Technology* (Stockton Press, New York, 1989).
23. W. R. Engels, C. R. Preston, P. Thompson, W. B. Eggleston, *Focus* **8**, 6 (1986). Positions 23D and 48A had previously been tested together with both inserts on the same chromosome (*9*), but these were separated and tested individually here.
24. G. B. Gloor *et al.*, *Genetics* **135**, 81 (1993).
25. W. R. Engels, *Environ. Mutagen.* 1, 37 (1979).

26. We thank T. Barrows for technical help. Supported by U.S. Public Health Service grant GM30948, paper 3397, from the Laboratory of Genetics, University of Wisconsin.

22 November 1993; accepted 24 January 1994

Mutation of a *mutL* Homolog in Hereditary Colon Cancer

Nickolas Papadopoulos,* Nicholas C. Nicolaides,* Ying-Fei Wei, Steven M. Ruben, Kenneth C. Carter, Craig A. Rosen, William A. Haseltine, Robert D. Fleischmann, Claire M. Fraser, Mark D. Adams, J. Craig Venter, Stanley R. Hamilton, Gloria M. Petersen, Patrice Watson, Henry T. Lynch, Päivi Peltomäki, Jukka-Pekka Mecklin, Albert de la Chapelle, Kenneth W. Kinzler,† Bert Vogelstein

Some cases of hereditary nonpolyposis colorectal cancer (HNPCC) are due to alterations in a *mutS*-related mismatch repair gene. A search of a large database of expressed sequence tags derived from random complementary DNA clones revealed three additional human mismatch repair genes, all related to the bacterial *mutL* gene. One of these genes (*hMLH1*) resides on chromosome 3p21, within 1 centimorgan of markers previously linked to cancer susceptibility in HNPCC kindreds. Mutations of *hMLH1* that would disrupt the gene product were identified in such kindreds, demonstrating that this gene is responsible for the disease. These results suggest that defects in any of several mismatch repair genes can cause HNPCC.

Hereditary nonpolyposis colorectal cancer is one of the most common genetic diseases of man, affecting as many as one in 200 individuals (*1*). Definitive evidence that the disease is inherited in a Mendelian fashion was provided when markers on chromosome 2p were found to segregate with disease in two large HNPCC kindreds (*2*). A further clue to pathogenesis emerged from the observation that a subset of sporadic colorectal cancers and most colorectal cancers from HNPCC patients exhibit widespread alterations of short, repeated sequences (microsatellites) distributed

throughout the genome (*3*). Experiments with tumor cell extracts suggested that these alterations are associated with a defect in an early step of mismatch repair (*4*). In bacteria and yeast, the *mutS* and *mutL* genes participate in early steps of mismatch repair, and inactivation of these genes results in microsatellite instability similar to that observed in tumor cells (*5*, *6*). It was therefore hypothesized that homologs of *mutS* or *mutL* might be responsible for HNPCC in humans (*4*, *6*, *7*). In support of this hypothesis, a human *mutS* homolog from chromosome 2p (*hMSH2*) was recently cloned (*8*, *9*), and mutations of this gene identified in the germline of HNPCC patients (*9*).

To determine whether mutations of other mismatch repair genes could cause HNPCC, we searched for additional homologs of bacterial and yeast mutator genes by scanning a database of human genes identified by the expressed sequence tag (EST) method (*10*). Three previously undescribed genes were identified that had significant similarity to *mutL* at their 5′ ends (predicted products shown in Fig. 1A). The first had strongest similarity to the yeast *mutL* gene *MLH1* (*11*) and was therefore termed h*MLH1*. The other two genes had slightly greater similarity to the yeast *mutL* homolog *PMS1* (*12*) and were denoted *hPMS1* and *hPMS2*, respectively.

The sequences of these three clones were used to design primers for polymerase chain reaction (PCR) assays with human genomic DNA as template. Somatic cell hybridization studies localized the *hMLH1*, *hPMS1*, and *hPMS2* genes to chromosomes 3, 2, and 7, respectively (*13*). The *hMLH1* gene was of particular interest because markers on chromosome 3p21 have been linked to HNPCC in several families (*14*, *15*). To determine the precise chromosomal location of *hMLH1*, a genomic DNA clone was used for fluorescence in situ hybridization (FISH) to human chromosome spreads

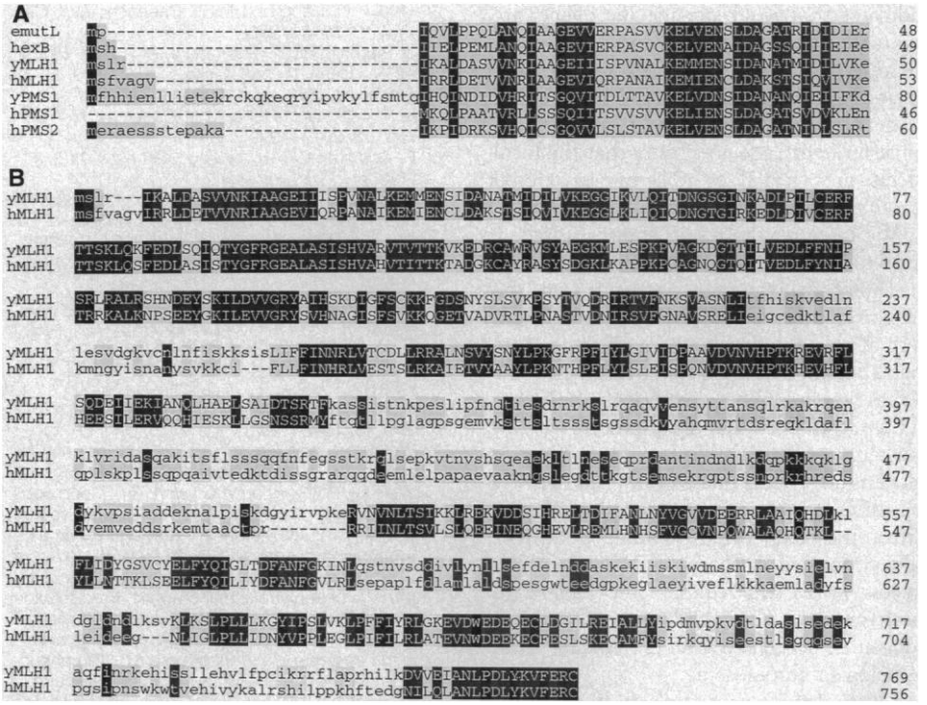

Fig. 1. Sequence comparison of bacterial, yeast, and human *mutL* gene products. (**A**) The predicted amino-terminal sequences of *mutL*-related gene products from *Escherichia coli* (emutL), *Streptococcus pneumoniae* (hexB) (*31*), *Saccharomyces cerevisiae* (yMLH1, yPMS1) (*11*, *12*), and human (hMLH1, hPMS1, hPMS2). The human sequences are those found in the EST screen (*10*). (**B**) The entire ORF of *hMLH1*, compared to its yeast counterpart. There is a polymorphism at codon 219 of *hMLH1*: Of 18 alleles tested, 63% encode Ile (ATC) and 37% encode Val (GTC). The sequence of the human gene has been deposited in GenBank (accession number U07418). The predicted amino acid sequences of yeast *MLH1* and *hMLH1* were analyzed with the MACAW (version 1.0) program. Amino acids in conserved blocks are capitalized and shaded based on the mean of their pairwise scores (*32*).

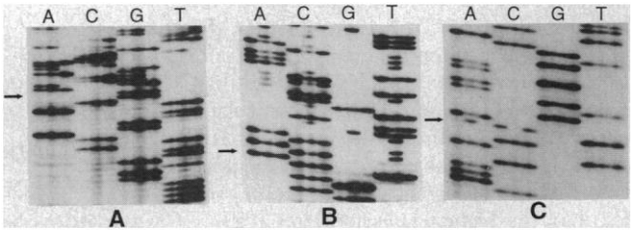

Fig. 2. Mutations of *hMLH1*, detected by sequence analysis of RT-PCR products. (**A**) A deletion beginning at codon 578 (arrow) in HNPCC kindreds F2 and F10 (left two lanes). (**B**) A 4-bp deletion at codon 727 (arrow) in HNPCC kindred L7 (middle lane). (**C**) A transversion of C to A at codon 252 (arrow) in H6 tumor cells (middle lane). In (A) and (B), RT-PCR products were generated from RNA of lymphoblastoid cells (*20*). In (C), genomic DNA was used for PCR (*33*). Sense primers were used for all sequencing reactions shown. The dideoxyadenosine triphosphate mixes for each reaction were loaded in adjacent lanes to facilitate comparison, as were those for the other nucleotides. The right lane in (A) and the outside lanes in (B) and (C) show the normal *hMLH1* sequence.

N. Papadopoulos, N. C. Nicolaides, K. W. Kinzler, B. Vogelstein, The Johns Hopkins Oncology Center, Baltimore, MD 21231, USA.
Y.-F. Wei, S. M. Ruben, K. C. Carter, C. A. Rosen, W. A. Haseltine, Human Genome Sciences, Inc., Suite 300, 9620 Medical Center Drive, Rockville, MD 20850–3338, USA.
R. D. Fleischmann, C. M. Fraser, M. D. Adams, J. C. Venter, The Institute for Genomic Research, 932 Clopper Road, Gaithersburg, MD 20878, USA.
S. R. Hamilton, Departments of Pathology and Oncology, The Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.
G. M. Petersen, Department of Epidemiology, The Johns Hopkins University School of Public Health and Hygiene, Baltimore, MD 21205, USA.
P. Watson and H. T. Lynch, Departments of Preventive Medicine and Public Health, Creighton University School of Medicine, Omaha, NE 68178, USA.
P. Peltomäki, J.-P. Mecklin, A. de la Chapelle, Department of Medical Genetics, University of Helsinki, 00290 Helsinki, Finland.

*These authors contributed equally to this work.
†To whom correspondence should be addressed.

(*16*). Of ~50 spreads analyzed, 40 had a doublet signal (characteristic of genuine hybridization) on at least one chromosome 3 homolog. Doublet signals were not detected on any other chromosomes. Detailed analysis of 17 individual spreads, counterstained to reveal banding, indicated that the *hMLH1* gene was located within band 3p21.3.

To evaluate the relationship between *hMLH1* and the HNPCC locus, we physically mapped the region. Detailed linkage analysis of HNPCC families has shown that the responsible locus is centered at markers D3S1611 and D3S1277, and is delimited by a 1-centimorgan (cM) interval bordered by markers D3S1298 and D3S1561 (*15*). No recombinations have been observed between markers D3S1611 or D3S1277 and cancer susceptibility. A yeast artificial chromosome (YAC) clone was identified that contained both *hMLH1* (*17*) and marker D3S1611, indicating that the gene is within 1 cM of the HNPCC locus. Analysis of six genomic P1 clones containing parts of *hMLH1* (*18*) showed

Table 1. Summary of mutations in *hMLH1*. Mutations were detected by direct sequencing of RT-PCR products. Each mutation was confirmed by replicate PCR analyses and by cloning and sequencing of the abnormal PCR products. The mutations in the kindreds were heterozygous; that is, lymphocytes from affected individuals contained one wild-type allele and one mutant allele. The wild-type allele was not detected in the H6 colorectal tumor cell line.

Sample	Codon	cDNA change	Predicted coding change
		Kindred	
F2, F3, F6, F8, F10, F11, F59	578 to 632	165-bp deletion	In-frame deletion
L7	727 and 728	4-bp deletion (TCACACATTC to TCATTCT)	Frameshift and substitution of new amino acids
L2516	755 and 756	4-bp insertion (GTGTTAA to GTGTTTGTTAA)	Extension of COOH-terminus
RA	347	371-bp deletion	Frameshift and truncation
		Cell line	
H6	252	Transversion (TCA to TAA)	Ser to stop codon

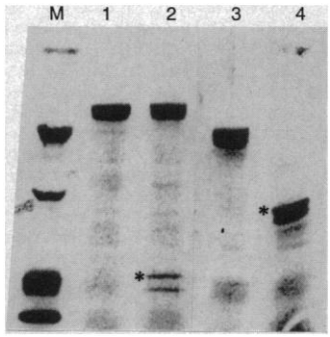

Fig. 3. Abnormal polypeptides produced from mutant *hMLH1* genes. RT-PCR products were synthesized with RNA templates from normal lymphoblasts (lanes 1 and 3), lymphoblasts of an affected member of kindred RA (lane 2), or from H6 colorectal tumor cells (lane 4). The products were transcribed and translated in vitro (*25, 27*), and the proteins separated on an SDS-polyacrylamide gel that was then subjected to autoradiography. In lanes 1 and 2, the RT-PCR product included codons 326 to 729, and the normal protein migrated at 48 kD. In lanes 3 and 4, the RT-PCR product included codons 1 to 394, and the normal protein migrated at 45 kD. The doublets result from internal initiation in the in vitro translated proteins. The truncated protein bands (asterisks) in lane 2 (18 kD) and in lane 4 (27 kD) are fainter than the normal protein bands because less [^{35}S]methionine is incorporated. Molecular size markers in the M lane correspond (top to bottom) to 68, 43, 29, 18, and 14 kD.

that the entire *hMLH1* gene was <85 kb and was totally contained within one of the P1 clones (P1417). Through coincidence, the randomly derived marker D3S1611 was found to be located in an intron of *hMLH1*.

The *hMLH1* gene thus became an excellent candidate for the HNPCC gene on chromosome 3p21. The sequence of the *hMLH1* gene was derived from four overlapping cDNA clones and numerous RT (reverse transcriptase)-PCR products derived from cellular RNA (Fig. 1). An open reading frame (ORF) began 42 nucleotides (nts) downstream of the 5′ end of the cDNA, and continued for 2268 nts. The methionine initiating this ORF was in a sequence context compatible with efficient translation and was preceded by an in-frame stop codon. The similarity of this ORF to yeast MLH1 was apparent throughout its length (34% amino acid identity; Fig. 1B). Several cDNA clones were apparently full length, as they hybridized to an mRNA of the same size (2.7 kb) as the cDNA insert (*19*). In addition, EST sequencing, Northern (RNA) blots, and RT-PCR analyses showed that the *hMLH1* gene is expressed in many tissues including colon, lymphocytes, breast, lung, spleen, testes, prostate, thyroid, gall bladder, and heart, in keeping with its presumptive housekeeping function.

To search for mutations in *hMLH1*, we performed RT-PCR analyses of lymphoblastoid cell RNA and directly sequenced the coding region of this gene in 10 HNPCC kindreds linked to 3p markers (*20, 21*). The intron-exon borders of selected exons were also determined and genomic PCR products analyzed to confirm the RT-PCR results. PCR products harboring suspected mutations were then cloned and sequenced to validate the results of the direct sequencing (*22*). Remarkably, affected individuals from seven Finnish kindreds all exhibited a heterozygous deletion of codons 578 to 632 (Fig. 2A). The derivation of five of these seven kindreds could be traced to a common ancestor (*15*), and the presence of the same presumptive defect in two other kindreds supported a "founder effect" responsible for many cases of HNPCC in the Finnish population. The deletion was not found in the cDNA of 61 other individuals tested (derived from normal or cancer patients, and including 54 lymphoblastoid lines and seven tumor cell lines). The genomic sequences surrounding codons 578 to 632 were determined by sequence analysis of the P1 clones (*23*) and the PCR products of genomic DNA (*24*). Codons 578 to 632 were found to constitute a single exon that was deleted from one allele in the kindreds described above. This exon encodes several highly conserved amino acids found at identical positions in yeast MLH1 (Fig. 1B).

In a second 3p-linked family (L7), we observed a 4-nt deletion beginning at the first position of codon 727 (Fig. 2B). This deletion produces a frameshift with a new stop codon located 166 nts downstream. As a result, the COOH-terminal 29 amino acids of hMLH1 are substituted with 53 different amino acids, some encoded by nucleotides normally in the 3′ untranslated region. Kindred L2516 displayed a different mutation, a 4-nt insertion between codons 755 and 756. This insertion results in a frameshift and extension of the ORF to include 99 nts downstream of the normal stop codon. The mutations in both kindreds L7 and L2516 were therefore predicted to alter the COOH-terminus of hMLH1. The extreme COOH-terminus contains 13 amino acids that are perfectly conserved between yeast MLH1 and hMLH1 (Fig. 1B), suggesting an important functional role for this region of the protein. Thus, *hMLH1* mutations were identified in nine of ten 3p-linked families (Table 1).

To examine HNPCC kindreds in which too few members were available for linkage studies, we used a screening technique that detects deletions, insertions, frameshifts, or nonsense mutations that alter the size of the gene product (*25*). We used RNA from lymphoblastoid cells of patients from 18 kindreds to amplify two products, one extending from codons 1 to 394 and the other from codons 326 to 729. Each of these kindreds met the International Collaborative Group criteria for HNPCC [three family members affected with colorectal cancer, at least two generations involved, at least one patient under 50 years of age at diagnosis (*26*)]. The PCR products were then transcribed and translated in vitro (*27*). All samples generated a polypeptide of the expected size, but a shorter polypeptide was also generated in one case (Fig. 3, lane 2).

Sequence analysis of the PCR product revealed a complex 371-nt deletion beginning at the first position of codon 347. This alteration was present in heterozygous form and resulted in a frameshift and a new stop codon, thus explaining the truncated polypeptide.

The same technique was used to examine four colorectal tumor cell lines manifesting microsatellite instability. One of the four (cell line H6) (*4*) generated only a truncated polypeptide (Fig. 3, lane 4). Sequence analysis of the cDNA revealed a C to A transversion at codon 252, resulting in the substitution of a stop codon for Ser (Table 1). No band at the normal C position was identified in the cDNA or genomic DNA from the H6 cells (Fig. 2C), indicating that these cells were devoid of a wild-type *hMLH1* allele.

Our results strongly suggest that HNPCC is in general associated with hereditary defects in mismatch repair genes. We estimate that mutations of *hMLH1* and *hMSH2* account for the majority of such defects, as most HNPCC families large enough for linkage analysis have been associated with markers linked to these two genes (*2, 14, 15, 28*). However, some HNPCC families are unlinked to these loci and may harbor mutations of other *mut* genes, such as *hPMS1* or *hPMS2*.

The results with the H6 tumor cell line are consistent with a model for HNPCC in which a second mutation is required for tumorigenesis. Lymphoblasts from HNPCC patients are repair proficient at the biochemical level (*4*), but tumors manifesting microsatellite instability are almost completely devoid of repair activity in the same assay. In particular, extracts of H6 cells have virtually no mismatch repair activity in vitro (*4*), an observation explained here by the total absence of the wild-type *hMLH1* allele. Two other tumors with microsatellite instability have been found to contain no wild-type *hMSH2* alleles (*9, 28*). Thus, inactivation of both copies of the relevant mismatch repair gene is likely to be a critical event in tumor formation in HNPCC patients. Once tumor cells sustain an alteration of the second allele, mutations rapidly accumulate, leading to the accelerated development of cancer in these families (*29*). This hypothesis is formally similar to Knudson's "two-hit" model for the inactivation of tumor suppressor genes in other cancer predisposition syndromes (*30*). The tumor specificity associated with mutations in these ubiquitously expressed mismatch repair genes remains a puzzle.

Finally, the discovery of the *hMLH1* gene illustrates the power of the random cDNA sequencing approach, which is likely to become increasingly useful as more expressed genes are identified and as more sophisticated programs for discerning functional homologies are developed.

REFERENCES AND NOTES

1. H. T. Lynch *et al.*, *Gastroenterology* **104**, 1535 (1993).
2. P. Peltomäki *et al.*, *Science* **260**, 810 (1993).
3. Y. M. Ionov, A. Peinado, S. Malkhosyan, D. Shibata, M. Perucho, *Nature* **363**, 558 (1993); S. N. Thibodeau, G. Bren, D. Schaid, *Science* **260**, 816 (1993); L. A. Aaltonen *et al.*, *ibid*, p. 812.
4. R. Parsons *et al.*, *Cell* **75**, 1227 (1993).
5. G. Levinson and G. A. Gutman, *Nucleic Acids Res.* **15**, 5323 (1987).
6. M. Strand, T. A. Prolla, R. M. Liskay, T. D. Petes, *Nature* **365**, 274 (1993).
7. T. A. Kunkel, *ibid.*, p. 207.
8. R. Fishel *et al.*, *Cell* **75**, 1027 (1993).
9. F. S. Leach *et al.*, *ibid.*, p. 1215.
10. The EST method involves automated DNA sequence analysis of randomly selected cDNA clones [M. D. Adams *et al.*, *Science* **252**, 1651 (1991); *Nature* **355**, 632 (1992); *Nature Genet.* **4**, 256 (1993); *ibid.*, p. 373]. The sequences of the *mutL*-related gene products were searched against a database of 65,000 ESTs using blastn and tblastn [S. F. Altschul *et al.*, *J. Mol. Biol.* **215**, 403 (1990)]. Nine ESTs were matched. Seven of these (from gall bladder, fetal heart, thyroid, embryo, fetal lung fibroblast, and Jurkat T cell cDNA libraries) were found to be products of the same gene (*hMLH1*); the other two were products of distinct genes (*hPMS1* and *hPMS2*, from a T cell lymphoma and an endometrial tumor library, respectively).
11. T. A. Prolla, D.-M. Christie, R. M. Liskay, *Mol. Cell. Biol.* **14**, 407 (1994).
12. W. Kramer *et al.*, *J. Bacteriol.* **171**, 5339 (1989).
13. Primers for genomic PCR analysis of *hMLH1* were 5′-GTTGAACATCTAGACGTTTCC-3′ and 5′-CGGTTCACCACTGTCTCGTC-3′, resulting in a 94-bp PCR product. For *hPMS1*, the primers were 5′-AAAACAAGTCAGTGAATCCTC-3′ and 5′-ATAAAAACAGCATGAGTCTGG-3′, resulting in an 88-bp product. For *hPMS2*, the primers were 5′-GTTTCAGAGTTAAGCCTTCG-3′ and 5′-GCTTAAGGACTATGGAGTGG-3′, resulting in an 89-bp product. PCR was performed with the primers described above, and DNA templates from rodent-human hybrid cell lines were used to determine chromosome localization. The conditions used for PCR are outlined in (*21*).
14. A. Lindblom, P. Tannergard, B. Werelius, M. Nordenskjold, *Nature Genet.* **5**, 279 (1993).
15. M. Nyström-Lahti *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
16. We isolated genomic clones of *hMLH1* by screening a human λ FIXII library (Stratagene) with a labeled cDNA clone [T. Maniatis, E. F. Fritsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982)]. Chromosome spreads were prepared as in R. Verma and A. Babu [*Human Chromosomes: Manual of Basic Techniques* (Pergamon, New York, 1989), pp 115–116], except that bromodeoxyuridine was used to induce high-resolution bands. A genomic clone (λ *hMLH1-13*) was nick-translated with digoxigenin–deoxy-uridine 5′-triphosphate (Boehringer Mannheim), and FISH was performed as in C. V. Johnson *et al.* [*Methods Cell Biol.* **35**, 73 (1991)]. Chromosomes were counterstained simultaneously with 4,6-diamino-2-phenylindole and propidium iodide, producing a combination of C-bands and R-bands. Aligned images for precise mapping were obtained with a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters [C. V. Johnson *et al.*, *Genet. Anal. Tech. Appl.* **8**, 75 (1991)].
17. The CEPH A library (Research Genetics, Inc.) consisted of 21,000 YAC clones arrayed in a format that allowed facile screening and unambiguous identification of positive clones. A YAC clone (6E6) containing *hMLH1* was identified by PCR screening with the primers in (*13*).
18. A human genomic P1 library (Genome Systems) was screened by PCR with the primers (*13*) for the 5′ end of *hMLH1* and primers 5′-CCACTGTGTATAAAGGAATAC-3′ and 5′-GTGATATACAAAGTGTACCAAC-3′, resulting in an 86-bp PCR product, for the 3′ end of *hMLH1*. Six clones were identified that contained the 5′ end of *hMLH1*, and one clone (P1417) was found to contain both the 5′ and 3′ ends of *hMLH1*.
19. Y.-F. Wei *et al.*, unpublished data.
20. We generated cDNA from RNA of lymphoblastoid or tumor cells with Superscript II (Life Technologies). The cDNA was used as a template for PCR. Primers 5′-GCATCTAGACGTTTCCTTGGC-3′ and 5′-CATCCAAGCTTCTGTTCCCG-3′ allowed amplification of codons 1 to 394; 5′-GGGGTGCAGCAGCACATCG-3′ and 5′-GGAGGCAGAATGTGTGAGCG-3′ allowed amplification of codons 326 to 729; and 5′-TCCCAAAGAAGGACTTGCT-3′ and 5′-AGTATAAGTCTTAAGTGCTACC-3′ allowed amplification of codons 602 to 756 plus 128 nts of 3′ untranslated sequences. The PCR conditions were as in (*21*). The PCR products were sequenced with primers labeled at their 5′ end with T4 polynucleotide kinase, using SequiTherm Polymerase (Epicentre Technologies). The 3p linkage of the kindreds was ascertained in (*15*) and (*28*).
21. The conditions for all PCR analyses were 35 cycles at 95°C for 30 s, 52 to 58°C for 60 to 120 s, and 70°C for 60 to 120 s, in the buffer described in [D. Sidransky *et al.*, *Science* **252**, 706 (1991)].
22. The PCR products were cloned into T-tailed vectors [T. A. Holton and M. W. Graham, *Nucleic Acids Res.* **19**, 1156 (1991)] and sequenced with T7 polymerase (United States Biochemical). All primers used for sequencing are available from the authors upon request.
23. Primers were labeled with T4 polynucleotide kinase, and P1 clones were cycle-sequenced with SequiTherm Polymerase (Epicentre Technologies), as described by the manufacturer.
24. The primers used to amplify the exon containing codons 578 to 632 were 5′-TTTATGGTTTCTCACCTGCC-3′ and 5′-GTTATCTGCCCACCTCAGC-3′. The PCR product included 105 bp of intron sequence upstream of the exon and 117 bp downstream. No mutations in the PCR product were observed in the Finnish kindreds, so the deletion in the RNA was not due to a simple splice site mutation. However, the deletion was found only in transcripts of the allele segregating with the disease and was therefore *cis*-acting.
25. S. M. Powell *et al.*, *N. Engl. J. Med.* **329**, 1982 (1993). The primers for coupled transcription-translation of *hMLH1* were 5′-GGATCCTAATACGACTCACTATAGGGAGACCACCATGGCATCTAGACGTTTCCTTGGC-3′ and 5′-CATCCAAGCTTCTGTTCCCG-3′ for codons 1 to 394, and 5′-GGATCCTAATACGACTCACTATAGGGAGACCACCATGGGGGTGCAGCAGCACATCG-3′ and 5′-GGAGGCAGAATGTGTGAGCG-3′ for codons 326 to 729. The PCR products had signals for transcription by T7 RNA polymerase and for the initiation of translation at their 5′ ends.
26. H. F. A. Vasen, J.-P. Mecklin, P. Meera-Khan, H. T. Lynch, *Dis. Colon Rectum* **34**, 424 (1991).
27. The PCR products were used as templates in coupled transcription-translation reactions (*25*) containing 40 μCi of [^{35}S]methionine (ICN). Samples were diluted in sample buffer, boiled for 5 min, and analyzed by electrophoresis on SDS-polyacrylamide gels containing a gradient of 10 to 20% acrylamide. The gels were dried and subjected to autoradiography.
28. R. Parsons *et al.*, unpublished data.
29. J. R. Jass and S. M. Stewart, *Gut* **33**, 783 (1992); J. R. Jass, unpublished data.
30. A. G. Knudson, *Cancer Res.* **45**, 1437 (1985).
31. M. Prudhomme, B. Martin, V. Mejean, J.-P. Claverys, *J. Bacteriol.* **171**, 7196 (1989); J. A. Mankovitch, C. A. McIntyre, G. C. Walker, *ibid.*, p. 5325; H-C. T. Tsui, B. S. Mandavilli, M. E. Winkler, *Nucleic Acids Res.* **20**, 2379 (1992).

32. G. D. Schuler, S. F. Altschul, D. J. Lipman, *Proteins Struct. Funct. Genet.* **9**, 180 (1991).
33. Primers used for PCR of genomic DNA were 5′-TGATAGAAATTGGATGTGAGG-3′ and 5′-TGATGAAGAGTAAGAAGATGC-3′, encompassing codons 228 to 263.
34. We thank R. Fuldner for advice on PCR primer design; K. Klimek for cDNA sequencing; M. Lerner, J. Wasmuth, S. M. Naylor, D. J. Chen, D. Ledbetter, and the Coriell Institute for Medical Research for DNA from rodent-human hybrid cell lines; J. Lawrence and members of her laboratory for advice and assistance with FISH mapping; S. Booker, J. Cavalieri, E. Bronson, and S. Slominski for clinical coordination and specimen collection; M. Castro for oligonucleotide synthesis; and T. Gwiazda for preparing the manuscript. Supported by the Clayton Fund, the Folkhäsan Institute of Genetics, the Academy of Finland, the Sigrid Juselius Foundation, the Nebraska State Health Department of Cancer, the Council for Tobacco Research, and NIH grants CA35494, CA47527, and CA57435. B.V. is an American Cancer Society Research Professor.

22 February 1994; accepted 4 March 1994

Coatomer Interaction with Di-Lysine Endoplasmic Reticulum Retention Motifs

Pierre Cosson and François Letourneur*

Although signals for retention in the endoplasmic reticulum (ER) have been identified in the cytoplasmic domain of various ER-resident type I transmembrane proteins, the mechanisms responsible for ER retention are still unknown. Yeast and mammalian ER retention motifs interacted specifically in cell lysates with the coatomer, a polypeptide complex implicated in membrane traffic. Mutations that affect the ER retention capacity of the motifs also abolished binding of the coatomer. These results suggest a role for the coatomer in the retrieval of transmembrane proteins to the ER in both yeast and mammals.

After translocation in the ER, proteins destined to the exocytic pathway are believed to be transported to the cell surface by default, unless retained in subcellular compartments by specific signals (*1*). For instance, a short KDEL (*2*) motif at the COOH-terminal end of soluble proteins can cause efficient retention in the ER (*3*). In the case of type I transmembrane proteins, ER retention motifs have been described in their transmembrane domains (*4*). Cytoplasmic ER retention motifs have also been identified in various mammalian proteins resident in the ER (*5*). Essential elements of these motifs are two lysine residues located at positions −3 and −4 from the COOH-terminal end of the cytoplasmic domain (*6*). Residency in the ER that is mediated by KDEL and di-lysine motifs is believed to be achieved by continuous retrieval to the ER of tagged proteins from post-ER compartments (*7, 8*). The receptor for the KDEL signal has been cloned and is localized in post-ER compartments reached early in the transport process, where it can sort KDEL-tagged proteins and ensure their transport back to the ER. The molecular mechanisms for ER retrieval of di-lysine–tagged proteins are unknown.

WBP1 is an ER-resident transmembrane protein from yeast that is a component of the yeast *N*-oligosaccharyl transferase complex (*9*). Its cytoplasmic domain contains a di-lysine motif (Fig. 1A) that acts as an ER retention signal in yeast (*10*). To test whether the yeast di-lysine ER retention signal also functions in mammalian cells, we transferred the cytoplasmic domain of WBP1 to the COOH-terminus of the Tac antigen (T-WBP1), a mammalian type I membrane protein normally expressed at the cell surface (*11*) (Fig. 1B). In transiently transfected mammalian cells, T-WBP1 was not detected at the cell surface of unpermeabilized cells (Fig. 1C) and appeared to be localized in the ER in permeabilized cells (Fig. 1E). Pulse-chase analysis confirmed that T-WBP1 was resident in the ER because no carbohydrate processing was detected during a 6-hour period (Fig. 1F). As observed for similar mammalian ER retention motifs (*5, 6*), a mutant of T-WBP1 in which both lysine residues were changed to serine (T-WBP1-SS) (Fig. 1A) was transported normally to the cell surface (Fig. 1, D and F). Mutants with the two lysine residues changed to arginine (T-WBP1-RR) or with four serine residues added at the COOH-terminus (T-WBP1+4S) (Fig. 1A) also lost ER retention and were transported to the cell surface (*12*). Thus, the yeast WBP1 di-lysine ER retention motif is functional in mammalian cells, which suggests conservation of the ER retention machinery between yeast and mammals.

To identify putative cytoplasmic receptors for the di-lysine ER retention motif, we constructed a chimeric protein containing the cytoplasmic domain of WBP1 fused to the COOH-terminus of glutathione-S-transferase (GST-WBP1). The fusion protein was expressed in bacteria and purified on glutathione-Sepharose beads. When yeast lysates were applied to the GST-WBP1 beads, we observed the specific binding of a 170-kD protein and of a set of proteins around 100 to 110 kD (Fig. 2A). These proteins did not bind to GST-

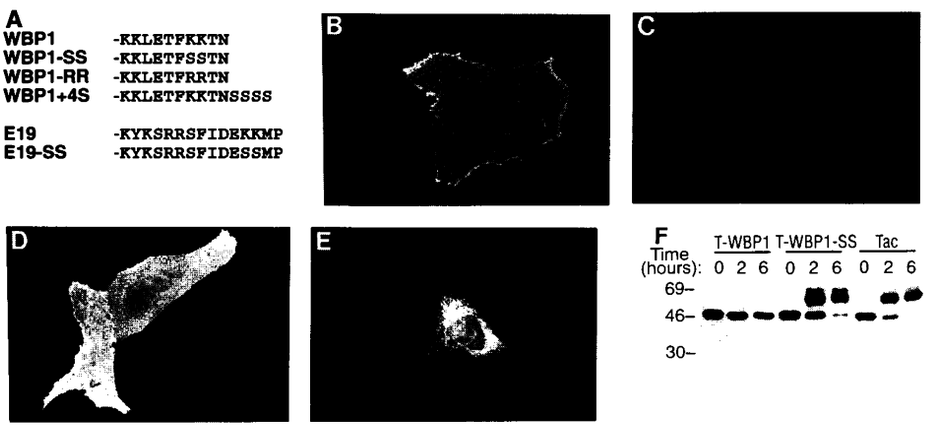

Fig. 1. Retention of Tac chimeras in the ER. (**A**) Sequences of the COOH-terminal domains of Tac and GST chimeric proteins (*2, 23*). (**B** through **E**) Intracellular localization of Tac chimeras by immunofluorescence microscopy. Intact CV cells transfected with Tac (B), T-WBP1 (C), or T-WBP1-SS (D) were stained with an antibody to Tac and a rhodamine-conjugated donkey antiserum to mouse immunoglobin G. In (E), CV-1 cells transfected with T-WBP1 were fixed and permeabilized before the immunofluorescence staining. (**F**) Pulse-chase analysis of COS-1 cells transfected with Tac chimeras. COS-1 cells transfected with the indicated Tac chimeras were labeled for 15 min with [^{35}S]methionine and were then incubated for 0, 2, or 6 hours in the presence of unlabeled methionine. Sedimented cells were extracted with Triton X-100, and Tac proteins were isolated by immunoprecipitation before analysis by SDS-PAGE (10% gel) (*21*). The 7G7 antibody is directed against an epitope localized to the extracellular domain of the human Tac antigen (*24*). Protein standards are indicated at the left in kilodaltons. Pulse-chase and immunofluorescence experiments were done with CV-1 and COS-1 cells and gave identical results.

Basel Institute for Immunology, Grenzacherstrasse 487, Postfach CH-4005 Basel, Switzerland.

*To whom correspondence should be addressed.