Mutations of two *PMS* homologues in hereditary nonpolyposis colon cancer

Nicholas C. Nicolaides*, Nickolas Papadopoulos*, Bo Liu*, Ying-Fei Wei†, Kenneth C. Carter†, Steven M. Ruben†, Craig A. Rosen†, William A. Haseltine†, Robert D. Fleischmann‡, Claire M. Fraser‡, Mark D. Adams‡, J. Craig Venter‡, Malcolm G. Dunlop§, Stanley R. Hamilton‖, Gloria M. Petersen¶, Albert de la Chapelle#, Bert Vogelstein* & Kenneth W. Kinzler*††

* The Johns Hopkins Oncology Center, Baltimore, Maryland 21231, USA
† Human Genome Sciences, Inc., 9620 Medical Center Drive, Suite 300, Rockville, Maryland 20850-3338, USA
‡ The Institute for Genomic Research, 932 Clopper Road, Gaithersburg, Maryland 20878, USA
§ MRC Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, UK
‖ The Johns Hopkins University School of Medicine, Departments of Pathology and Oncology, Baltimore, Maryland 21205, USA
¶ Department of Epidemiology, The Johns Hopkins University School of Public Health and Hygiene, Baltimore, Maryland 21205, USA
# Department of Medical Genetics, University of Helsinki, 00290 Helsinki, Finland

HEREDITARY **nonpolyposis colorectal cancer (HNPCC) is one of man's commonest hereditary diseases[1]. Several studies have implicated a defect in DNA mismatch repair in the pathogenesis of this disease[2–8]. In particular, *hMSH2* and *hMLH1* homologues of the bacterial DNA mismatch repair genes *mutS* and *mutL*, respectively, were shown to be mutated in a subset of HNPCC cases[9–16]. Here we report the nucleotide sequence, chromosome localization and mutational analysis of *hPMS1* and *hPMS2*, two additional homologues of the prokaryotic *mutL* gene. Both *hPMS1* and *hPMS2* were found to be mutated in the germline of HNPCC patients. This doubles the number of genes implicated in HNPCC and may help explain the relatively high incidence of this disease.**

We previously screened a database of human genes identified by the expressed sequence tag (EST) method[17] and identified three ESTs with homology to bacterial and yeast *mutL* genes[16]. Two ESTs had greatest homology to the yeast *mutL*-homologue *yPMS1*, and therefore were designated *hPMS1* and *hPMS2*. Sequence analysis of the *hPMS1* complementary DNA clone identified an open reading frame (ORF) of 2,795 basepairs (bp) flanked on both sides by in-frame termination codons (Fig. 1*a*). If initiation occurred at the first methionine of the ORF, a 932 residue protein with 27% identity to *yPMS1* would be produced. Sequence analysis of the *hPMS2* clone identified a 2,586 bp ORF (Fig. 1*b*) without an upstream stop codon. Attempts to extend the 5′ end sequence using the rapid amplification of cDNA ends (RACE) technique[18], combined with genomic sequencing revealed a complex pattern suggesting alternative initiation sites (data not shown). Sequence analysis of the major 5′ products failed to reveal an upstream in-frame stop codon. However, the methionine at nucleotide 25 fit the context of a Kozak consensus sequence for translation initiation at 5 of 7 positions[19]. If initiation occurred at this methionine, an 862 residue protein with 32% identity to *yPMS1* would be produced. Comparison of the yeast and human PMS proteins showed that the greatest homology was within the amino-terminal segment, from codons 40 to 390, which was 30% identical between *yPMS1* and *hPMS1*, and 42% between *yPMS1* and *hPMS2* (Fig. 1*c*). This region contained several domains that are highly conserved among the *mutL*-related proteins, most notably the GFRGEAL domain which is perfectly conserved in E. coli and human *mutL* homologues[20]. A second region of significant homology was in the carboxyl terminus. This region was 22% identical between *yPMS1* and *hPMS1* and 47% identical between *yPMS1* and *hPMS2*, while very little homology was observed in this region between these proteins and the yeast *mutL*-homologue yMLH1 (not shown).

hPMS1 and *hPMS2* were previously localized to chromosomes 2 and 7, respectively[16]. To determine the precise location of these genes, genomic clones were used for fluorescent in situ hybridization (FISH) to human metaphase chromosome spreads. *hPMS1* was localized to chromosome 2q31-33 using a genomic P1 clone (1670) which contained its 5′ end (Fig. 2*a*, *b*, *c*). Likewise, the *hPMS2* gene was localized to chromosome 7p22 using a genomic P1 clone (2053) which contained the 3′ region of the *hPMS2* gene (Fig. 2*d*, *e*). Liskay and colleagues have also cloned a human *mutL* homologue which maps to chromosome 7 (R. M. Liskay, personal communication)[15]. Analysis with a variety of genomic clones indicates that *hPMS2* is a member of a subfamily which includes at least two related genes located on chromosome 7q.

To evaluate the role of *hPMS1* and *hPMS2* in HNPCC, we examined 40 patients with a family history of HNPCC. Eighteen were found to have mutations of *hMSH2* or *hMLH1* and therefore were not fully evaluated for *hPMS1* or *hPMS2* mutations[13,16,21]. To identify potential mutations in the other 22 samples, we employed an *in vitro* synthesized protein (IVSP) assay which detects deletions, insertions, frameshifts and nonsense mutations that result in an altered size of the protein product[22,23]. Three overlapping segments of *hPMS1* were amplified using reverse transcriptase polymerase chain reaction (RT-PCR) and the products were transcribed and translated *in vitro*. All samples generated a protein of the expected size, although a reproducible truncated product was detected in a segment containing codons 1 to 500 from patient C.W. The expected relative molecular mass of the wild-type protein from this segment was 55,000 (55 K) while the truncated protein product had an approximate molecular weight of 25 K. Sequence analysis of RT-PCR products from this region revealed an in-frame deletion that removed codons 195 to 233. Because this small deletion could not result in the 25 K truncated protein, we attempted to map the truncation by amplifying cDNA from C.W. using the same 5′ primer and nested 3′ primers. IVSP analysis revealed that the truncated polypeptide was present in samples translated from templates containing codons 1 to 369 and 1 to 290 but was absent in a template containing codons 1 to 214 (Fig. 3*a*). This placed the potential stop codon mutation between codons 214 and 290, in a region overlapping with the in-frame deletion. As it has been previously noted that exons containing nonsense codons can be selectively skipped[24], we determined the intron–exon boundaries in this region. Sequence analysis of an *hPMS1* genomic clone identified an exon which corresponded exactly to the boundaries of the deletion (codons 195–233). We then amplified the skipped exon from the genomic DNA of patient C.W. While no alterations of the splicing consensus sites were identified, a single C to T transition at codon 233, which converted a glutamine to stop codon, was identified (Fig. 3*b*, lane 2). This nonsense codon appeared to be responsible for the exon-skipping.

For mutational analysis of *hPMS2*, cDNA was amplified either as a full-length product or as two overlapping segments and used in the IVSP assay. In one case (G.C.), analysis of the full-length product identified an additional truncated protein of about 50 K (Fig. 4*a*, lane GC). Moreover, analysis of *hPMS2* in a tumour xenograft from this patient revealed the 50 K truncated

†† To whom correspondence should be addressed

FIG. 1 Structure of human *hPMS1* and *hPMS2*. The cDNA sequence and predicted amino acid sequence of the protein encoded by *hPMS1* (*a*) and *hPMS2* (*b*) are shown. *c* (shown on page 78), Alignment of the predicted amino-acid sequences of the *S. cerevisiae* PMS1 (*yPMS1*)[26], the human PMS1 (*hPMS1*) and the human PMS2 (*hPMS2*) proteins using MACAW (version 1.0) program. Amino acids in conserved blocks are capitalized and shaded based on the mean of their pair wise scores[27]. The *hPMS1* and *hPMS2* nucleotide sequences have been deposited with Genbank (accession numbers U13695 and U13696, respectively).

a

```
   1                                                                                                            Met
   1   GG CAC GAG TGG CTG CTT GCG GCT AGT GGA TGG TAA TTG CCT GCC TCG CGC TAG CAG CAA GCT GCT CTG TTA AAA GCG AAA ATG

   2  Lys Gln Leu Pro Ala Ala Thr Val Arg Leu Leu Ser Ser Ser Gln Ile Ile Thr Ser Val Val Ser Val Val Lys Glu Leu Ile
  84  AAA CAA TTG CCT GCG GCA ACA GTT CGA CTC CTT TCA AGT TCT CAG ATC ATC ACT TCG GTG GTC AGT GTT GTA AAA GAG CTT ATT

  30  Glu Asn Ser Leu Asp Ala Gly Ala Thr Ser Val Asp Val Lys Leu Glu Asn Tyr Gly Phe Asp Lys Ile Glu Val Arg Asp Asn
 168  GAA AAC TCC TTG GAT GCT GGT GCC ACA AGC GTA GAT GTT AAA CTG GAG AAC TAT GGA TTT GAT AAA ATT GAG GTG CGA GAT AAC

  58  Gly Glu Gly Ile Lys Ala Val Asp Ala Pro Val Met Ala Met Lys Tyr Tyr Thr Ser Lys Ile Asn Ser His Glu Asp Leu Glu
 252  GGG GAG GGT ATC AAG GCT GTT GAT GCA CCT GTA ATG GCA ATG AAG TAC TAC ACC TCA AAA ATA AAT AGT CAT GAA GAT CTT GAA

  86  Asn Leu Thr Thr Tyr Gly Phe Arg Gly Glu Ala Leu Gly Ser Ile Cys Cys Ile Ala Glu Val Leu Ile Thr Thr Arg Thr Ala
 336  AAT TTG ACA ACT TAC GGT TTT CGT GGA GAA GCC TTG GGG TCA ATT TGT TGT ATA GCT GAG GTT TTA ATT ACA ACA AGA ACG GCT

 114  Ala Asp Asn Phe Ser Thr Gln Tyr Val Leu Asp Gly Ser Gly His Ile Leu Ser Gln Lys Pro Ser His Leu Gly Gln Gly Thr
 420  GCT GAT AAT TTT AGC ACC CAG TAT GTT TTA GAT GGC AGT GGC CAC ATA CTT TCT CAG AAA CCT TCA CAT CTT GGT CAA GGT ACA

 142  Thr Val Thr Ala Leu Arg Leu Phe Lys Asn Leu Pro Val Arg Lys Gln Phe Tyr Ser Thr Ala Lys Lys Cys Lys Asp Glu Ile
 504  ACT GTA ACT GCT TTA AGA TTA TTT AAG AAT CTA CCT GTA AGA AAG CAG TTT TAC TCA ACT GCA AAA AAA TGT AAA GAT GAA ATA

 170  Lys Lys Ile Gln Asp Leu Leu Met Ser Phe Gly Ile Leu Lys Pro Asp Leu Arg Ile Val Phe Val His Asn Lys Ala Val Ile
 588  AAA AAG ATC CAA GAT CTC CTC ATG AGC TTT GGT ATC CTT AAA CCT GAC TTA AGG ATT GTC TTT GTA CAT AAC AAG GCA GTT ATT

 198  Trp Gln Lys Ser Arg Val Ser Asp His Lys Met Ala Leu Met Ser Val Leu Gly Thr Ala Val Met Asn Asn Met Glu Ser Phe
 672  TGG CAG AAA AGC AGA GTA TCA GAT CAC AAG ATG GCT CTC ATG TCA GTT CTG GGG ACT GCT GTT ATG AAC AAT ATG GAA TCC TTT

 226  Gln Tyr His Ser Glu Glu Ser Gln Ile Tyr Leu Ser Gly Phe Leu Pro Lys Cys Asp Ala Asp His Ser Phe Thr Ser Leu Ser
 756  CAG TAC CAC TCT GAA GAA TCT CAG ATT TAT CTC AGT GGA TTT CTT CCA AAG TGT GAT GCA GAC CAC TCT TTC ACT AGT CTT TCA

 254  Thr Pro Glu Arg Ser Phe Ile Phe Ile Asn Ser Arg Pro Val His Gln Lys Asp Ile Leu Lys Leu Ile Arg His His Tyr Asn
 840  ACA CCA GAA AGA AGT TTC ATC TTC ATA AAC AGT CGA CCA GTA CAT CAA AAA GAT ATC TTA AAG TTA ATC CGA CAT CAT TAC AAT

 282  Leu Lys Cys Leu Lys Glu Ser Thr Arg Leu Tyr Pro Val Phe Phe Leu Lys Ile Asp Val Pro Thr Ala Asp Val Asp Val Asn
 924  CTG AAA TGC CTA AAG GAA TCT ACT CGT TTG TAT CCT GTT TTC TTT CTG AAA ATC GAT GTT CCT ACA GCT GAT GTT GAT GTA AAT

 310  Leu Thr Pro Asp Lys Ser Gln Val Leu Leu Gln Asn Lys Glu Ser Val Leu Ile Ala Leu Glu Asn Leu Met Thr Thr Cys Tyr
1008  TTA ACA CCA GAT AAA AGC CAA GTA TTA TTA CAA AAT AAG GAA TCT GTT TTA ATT GCT CTT GAA AAT CTG ATG ACG ACT TGT TAT

 338  Gly Pro Leu Pro Ser Thr Asn Ser Tyr Glu Asn Asn Lys Thr Asp Val Ser Ala Ala Asp Ile Val Leu Ser Lys Thr Ala Glu
1092  GGA CCA TTA CCT AGT ACA AAT TCT TAT GAA AAT AAT AAA ACA GAT GTT TCC GCA GCT GAC ATC GTT CTT AGT AAA ACA GCA GAA

 366  Thr Asp Val Leu Phe Asn Lys Val Glu Ser Ser Gly Lys Asn Tyr Ser Asn Val Asp Thr Ser Val Ile Pro Phe Gln Asn Asp
1176  ACA GAT GTG CTT TTT AAT AAA GTG GAA TCA TCT GGA AAG AAT TAT TCA AAT GTT GAT ACT TCA GTC ATT CCA TTC CAA AAT GAT

 394  Met His Asn Asp Glu Ser Gly Lys Asn Thr Asp Asp Cys Leu Asn His Gln Ile Ser Ile Gly Asp Phe Gly Tyr Gly His Cys
1260  ATG CAT AAT GAT GAA TCT GGA AAA AAC ACT GAT GAT TGT TTA AAT CAC CAG ATA AGT ATT GGT GAC TTT GGT TAT GGT CAT TGT

 422  Ser Ser Glu Ile Ser Asn Ile Asp Lys Asn Thr Lys Asn Ala Phe Gln Asp Ile Ser Met Ser Asn Val Ser Trp Glu Asn Ser
1344  AGT AGT GAA ATT TCT AAC ATT GAT AAA AAC ACT AAG AAT GCA TTT CAG GAC ATT TCA ATG AGT AAT GTA TCA TGG GAG AAC TCT

 450  Gln Thr Glu Tyr Ser Lys Thr Cys Phe Ile Ser Ser Val Lys His Thr Gln Ser Glu Asn Gly Asn Lys Asp His Ile Asp Glu
1428  CAG ACG GAA TAT AGT AAA ACT TGT TTT ATA AGT TCC GTT AAG CAC ACC CAG TCA GAA AAT GGC AAT AAA GAC CAT ATA GAT GAG

 478  Ser Gly Glu Asn Glu Glu Glu Ala Gly Leu Glu Asn Ser Ser Glu Ile Ser Ala Asp Glu Trp Ser Arg Gly Asn Ile Leu Lys
1512  AGT GGG GAA AAT GAG GAA GAA GCA GGT CTT GAA AAC TCT TCG GAA ATT TCT GCA GAT GAG TGG AGC AGG GGA AAT ATA CTT AAA

 506  Asn Ser Val Gly Glu Asn Ile Glu Pro Val Lys Ile Leu Val Pro Glu Lys Ser Leu Pro Cys Lys Val Ser Asn Asn Asn Tyr
1596  AAT TCA GTG GGA GAG AAT ATT GAA CCT GTG AAA ATT TTA GTG CCT GAA AAA AGT TTA CCA TGT AAA GTA AGT AAT AAT AAT TAT

 534  Pro Ile Pro Glu Gln Met Asn Leu Asn Glu Asp Ser Cys Asn Lys Lys Ser Asn Val Ile Asp Asn Lys Ser Gly Lys Val Thr
1680  CCA ATC CCT GAA CAA ATG AAT CTT AAT GAA GAT TCA TGT AAC AAA AAA TCA AAT GTA ATA GAT AAT AAA TCT GGA AAA GTT ACA

 562  Ala Tyr Asp Leu Leu Ser Asn Arg Val Ile Lys Lys Pro Met Ser Ala Ser Ala Leu Phe Val Gln Asp His Arg Pro Gln Phe
1764  GCT TAT GAT TTA CTT AGC AAT CGA GTA ATC AAG AAA CCC ATG TCA GCA AGT GCT CTT TTT GTT CAA GAT CAT CGT CCT CAG TTT

 590  Leu Ile Glu Asn Pro Lys Thr Ser Leu Glu Asp Ala Thr Leu Gln Ile Glu Glu Leu Trp Lys Thr Leu Ser Glu Glu Glu Lys
1848  CTC ATA GAA AAT CCT AAG ACT AGT TTA GAG GAT GCA ACA CTA CAA ATT GAA GAA CTG TGG AAG ACA TTG AGT GAA GAG GAA AAA

 618  Leu Lys Tyr Glu Glu Lys Ala Thr Lys Asp Leu Glu Arg Tyr Asn Ser Gln Met Lys Arg Ala Ile Glu Gln Glu Ser Gln Met
1932  CTG AAA TAT GAA GAG AAG GCT ACT AAA GAC TTG GAA CGA TAC AAT AGT CAA ATG AAG AGA GCC ATT GAA CAG GAG TCA CAA ATG

 646  Ser Leu Lys Asp Gly Arg Lys Lys Ile Lys Pro Thr Ser Ala Trp Asn Leu Ala Gln Lys His Lys Leu Lys Thr Ser Leu Ser
2016  TCA CTA AAA GAT GGC AGA AAA AAG ATA AAA CCC ACC AGC GCA TGG AAT TTG GCC CAG AAG CAC AAG TTA AAA ACC TCA TTA TCT

 674  Asn Gln Pro Lys Leu Asp Glu Leu Leu Gln Ser Gln Ile Glu Lys Arg Arg Ser Gln Asn Ile Lys Met Val Gln Ile Pro Phe
2100  AAT CAA CCA AAA CTT GAT GAA CTC CTT CAG TCC CAA ATT GAA AAA AGA AGG AGT CAA AAT ATT AAA ATG GTA CAG ATC CCC TTT

 702  Ser Met Lys Asn Leu Lys Ile Asn Phe Lys Lys Gln Asn Lys Val Asp Leu Glu Glu Lys Asp Glu Pro Cys Leu Ile His Asn
2184  TCT ATG AAA AAC TTA AAA ATA AAT TTT AAG AAA CAA AAC AAA GTT GAC TTA GAA GAG AAG GAT GAA CCT TGC TTG ATC CAC AAT

 730  Leu Arg Phe Pro Asp Ala Trp Leu Met Thr Ser Lys Thr Glu Val Met Leu Leu Asn Pro Tyr Arg Val Glu Glu Ala Leu Leu
2268  CTC AGG TTT CCT GAT GCA TGG CTA ATG ACA TCC AAA ACA GAG GTA ATG TTA TTA AAT CCA TAT AGA GTA GAA GAA GCC CTG CTA

 758  Phe Lys Arg Leu Leu Glu Asn His Lys Leu Pro Ala Glu Pro Leu Glu Lys Pro Ile Met Leu Thr Glu Ser Leu Phe Asn Gly
2352  TTT AAA AGA CTT CTT GAG AAT CAT AAA CTT CCT GCA GAG CCA CTG GAA AAG CCA ATT ATG TTA ACA GAG AGT CTT TTT AAT GGA

 786  Ser His Tyr Leu Asp Val Leu Tyr Lys Met Thr Ala Asp Asp Gln Arg Tyr Ser Gly Ser Thr Tyr Leu Ser Asp Pro Arg Leu
2436  TCT CAT TAT TTA GAC GTT TTA TAT AAA ATG ACA GCA GAT GAC CAA AGA TAC AGT GGA TCA ACT TAC CTG TCT GAT CCT CGT CTT

 814  Thr Ala Asn Gly Phe Lys Ile Lys Leu Ile Pro Gly Val Ser Ile Thr Glu Asn Tyr Leu Glu Ile Glu Gly Met Ala Asn Cys
2520  ACA GCG AAT GGT TTC AAG ATA AAA TTG ATA CCA GGA GTT TCA ATT ACT GAA AAT TAC TTG GAA ATA GAA GGA ATG GCT AAT TGT

 842  Leu Pro Phe Tyr Gly Val Ala Asp Leu Lys Glu Ile Leu Asn Ala Ile Leu Asn Arg Asn Ala Lys Glu Val Tyr Glu Cys Arg
2604  CTC CCA TTC TAT GGA GTA GCA GAT TTA AAA GAA ATT CTT AAT GCT ATA TTA AAC AGA AAT GCA AAG GAA GTT TAT GAA TGT AGA

 870  Pro Arg Lys Val Ile Ser Tyr Leu Glu Gly Glu Ala Val Arg Leu Ser Arg Gln Leu Pro Met Tyr Leu Ser Lys Glu Asp Ile
2688  CCT CGC AAA GTG ATA AGT TAT TTA GAG GGA GAA GCA GTG CGT CTA TCC AGA CAA TTA CCC ATG TAC TTA TCA AAA GAG GAC ATC

 898  Gln Asp Ile Ile Tyr Arg Met Lys His Gln Phe Gly Asn Glu Ile Lys Glu Cys Val His Gly Arg Pro Phe Phe His His Leu
2772  CAA GAC ATT ATC TAC AGA ATG AAG CAC CAG TTT GGA AAT GAA ATT AAA GAG TGT GTT CAT GGT CGC CCA TTT TTT CAT CAT TTA

 926  Thr Tyr Leu Pro Glu Thr Thr
2856  ACC TAT CTT CCA GAA ACT ACA TGA TTA AAT ATG TTT AAG AAG ATT AGT TAC CAT TGA AAT TGG TTC TGT CAT AAA ACA GCA TGA

2940  GTC TGG TTT TAA ATT ATC TTT GTA TTA TGT GTC ACA TGG TTA TTT TTT AAA TGA GGA TTC ACT GAC TTG TTT TTA TAT TGA AAA

3024  AAG TTC CAC GTA TTG TAG AAA ACG TAA ATA AAC TAA TAA C
```

b

```
   1                                   Met Glu Arg Ala Glu Ser Ser Ser Thr Glu Pro Ala Lys Ala Ile Lys Pro Ile Asp Arg
   1  CGA GGC GGA TCG GGT GTT GCA TCC ATG GAG CGA GCT GAG AGC TCG AGT ACA GAA CCT GCT AAG GCC ATC AAA CCT ATT GAT CGG

  21  Lys Ser Val His Gln Ile Cys Ser Gly Gln Val Val Leu Ser Leu Ser Thr Ala Val Lys Glu Leu Val Glu Asn Ser Leu Asp
  85  AAG TCA GTC CAT CAG ATT TGC TCT GGG CAG GTG GTA CTG AGT CTA AGC ACT GCG GTA AAG GAG TTA GTA GAA AAC AGT CTG GAT

  49  Ala Gly Ala Thr Asn Ile Asp Leu Lys Leu Lys Asp Tyr Gly Val Asp Leu Ile Glu Val Ser Asp Asn Gly Cys Gly Val Glu
 169  GCT GGT GCC ACT AAT ATT GAT CTA AAG CTT AAG GAC TAT GGA GTG GAT CTT ATT GAA GTT TCA GAC AAT GGA TGT GGG GTA GAA

  77  Glu Glu Asn Phe Glu Gly Leu Thr Leu Lys His His Thr Ser Lys Ile Gln Glu Phe Ala Asp Leu Thr Gln Val Glu Thr Phe
 253  GAA GAA AAC TTC GAA GGC TTA ACT CTG AAA CAT CAC ACA TCT AAG ATT CAA GAG TTT GCC GAC CTA ACT CAG GTT GAA ACT TTT

 105  Gly Phe Arg Gly Glu Ala Leu Ser Ser Leu Cys Ala Leu Ser Asp Val Thr Ile Ser Thr Cys His Ala Ser Ala Lys Val Gly
 337  GGC TTT CGG GGG GAA GCT CTG AGC TCA CTT TGT GCA CTG AGC GAT GTC ACC ATT TCT ACC TGC CAC GCA TCG GCG AAG GTT GGA

 133  Thr Arg Leu Met Phe Asp His Asn Gly Lys Ile Ile Gln Lys Thr Pro Tyr Pro Arg Pro Arg Gly Thr Thr Val Ser Val Gln
 421  ACT CGA CTG ATG TTT GAT CAC AAT GGG AAA ATT ATC CAG AAA ACC CCC TAC CCC CGC CCC AGA GGG ACC ACA GTC AGC GTG CAG

 161  Gln Leu Phe Ser Thr Leu Pro Val Arg His Lys Glu Phe Gln Arg Asn Ile Lys Lys Glu Tyr Ala Lys Met Val Gln Val Leu
 505  CAG TTA TTT TCC ACA CTA CCT GTG CGC CAT AAG GAA TTT CAA AGG AAT ATT AAG AAG GAG TAT GCC AAA ATG GTC CAG GTC TTA

 189  His Ala Tyr Cys Ile Ile Ser Ala Gly Ile Arg Val Ser Cys Thr Asn Gln Leu Gly Gln Gly Lys Arg Gln Pro Val Val Cys
 589  CAT GCA TAC TGT ATC ATT TCA GCA GGC ATC CGT GTA AGT TGC ACC AAT CAG CTT GGA CAA GGA AAA CGA CAG CCT GTG GTA TGC

 217  Thr Gly Gly Ser Pro Ser Ile Lys Glu Asn Ile Gly Ser Val Phe Gly Gln Lys Gln Leu Gln Ser Leu Ile Pro Phe Val Gln
 673  ACA GGT GGA AGC CCC AGC ATA AAG GAA AAT ATC GGC TCT GTG TTT GGG CAG AAG CAG TTG CAA AGC CTC ATT CCT TTT GTT CAG

 245  Leu Pro Pro Ser Asp Ser Val Cys Glu Glu Tyr Gly Leu Ser Cys Ser Asp Ala Leu His Asn Leu Phe Tyr Ile Ser Gly Phe
 757  CTG CCC CCT AGT GAC TCC GTG TGT GAA GAG TAC GGT TTG AGC TGT TCG GAT GCT CTG CAT AAT CTT TTT TAC ATC TCA GGT TTC

 273  Ile Ser Gln Cys Thr His Gly Val Gly Arg Ser Ser Thr Asp Arg Gln Phe Phe Phe Ile Asn Arg Arg Pro Cys Asp Pro Ala
 841  ATT TCA CAA TGC ACG CAT GGA GTT GGA AGG AGT TCA ACA GAC AGA CAG TTT TTC TTT ATC AAC CGG CGG CCT TGT GAC CCA GCA

 301  Lys Val Cys Arg Leu Val Asn Glu Val Tyr His Met Tyr Asn Arg His Gln Tyr Pro Phe Val Val Leu Asn Ile Ser Val Asp
 925  AAG GTC TGC AGA CTC GTG AAT GAG GTC TAC CAC ATG TAT AAT CGA CAC CAG TAT CCA TTT GTT GTT CTT AAC ATT TCT GTT GAT

 329  Ser Glu Cys Val Asp Ile Asn Val Thr Pro Asp Lys Arg Gln Ile Leu Leu Gln Glu Glu Lys Leu Leu Leu Ala Val Leu Lys
1009  TCA GAA TGC GTT GAT ATC AAT GTT ACT CCA GAT AAA AGG CAA ATT TTG CTA CAA GAG GAA AAG CTT TTG TTG GCA GTT TTA AAG

 357  Thr Ser Leu Ile Gly Met Phe Asp Ser Asp Val Asn Lys Leu Asn Val Ser Gln Gln Pro Leu Leu Asp Val Glu Gly Asn Leu
1093  ACC TCT TTG ATA GGA ATG TTT GAT AGT GAT GTC AAC AAG CTA AAT GTC AGT CAG CAG CCA CTG CTG GAT GTT GAA GGT AAC TTA

 385  Ile Lys Met His Ala Ala Asp Leu Glu Lys Pro Met Val Glu Lys Gln Asp Gln Ser Pro Ser Leu Arg Thr Gly Glu Glu Lys
1177  ATA AAA ATG CAT GCA GCG GAT TTG GAA AAG CCC ATG GTA GAA AAG CAG GAT CAA TCC CCT TCA TTA AGG ACT GGA GAA GAA AAA

 413  Lys Asp Val Ser Ile Ser Arg Leu Arg Glu Ala Phe Ser Leu Arg His Thr Thr Glu Asn Lys Pro His Ser Pro Lys Thr Pro
1261  AAA GAC GTG TCC ATT TCC AGA CTG CGA GAG GCC TTT TCT CTT CGT CAC ACA ACA GAG AAC AAG CCT CAC AGC CCA AAG ACT CCA

 441  Glu Pro Arg Arg Ser Pro Leu Gly Gln Lys Arg Gly Met Leu Ser Ser Ser Thr Ser Gly Ala Ile Ser Asp Lys Gly Val Leu
1345  GAA CCA AGA AGG AGC CCT CTA GGA CAG AAA AGG GGT ATG CTG TCT TCT AGC ACT TCA GGT GCC ATC TCT GAC AAA GGC GTC CTG

 469  Arg Pro Gln Lys Glu Ala Val Ser Ser Ser His Gly Pro Ser Asp Pro Thr Asp Arg Ala Glu Val Glu Lys Asp Ser Gly His
1429  AGA CCT CAG AAA GAG GCA GTG AGT TCC AGT CAC GGA CCC AGT GAC CCT ACG GAC AGA GCG GAG GTG GAG AAG GAC TCG GGG CAC

 497  Gly Ser Thr Ser Val Asp Ser Glu Gly Phe Ser Ile Pro Asp Thr Gly Ser His Cys Ser Ser Glu Tyr Ala Ala Ser Ser Pro
1513  GGC AGC ACT TCC GTG GAT TCT GAG GGG TTC AGC ATC CCA GAC ACG GGC AGT CAC TGC AGC AGC GAG TAT GCG GCC AGC TCC CCA

 525  Gly Asp Arg Gly Ser Gln Glu His Val Asp Ser Gln Glu Lys Ala Pro Glu Thr Asp Asp Ser Phe Ser Asp Val Asp Cys His
1597  GGG GAC AGG GGC TCG CAG GAA CAT GTG GAC TCT CAG GAG AAA GCG CCT GAA ACT GAC GAC TCT TTT TCA GAT GTG GAC TGC CAT

 553  Ser Asn Gln Glu Asp Thr Gly Cys Lys Phe Arg Val Leu Pro Gln Pro Thr Asn Leu Ala Thr Pro Asn Thr Lys Arg Phe Lys
1681  TCA AAC CAG GAA GAT ACC GGA TGT AAA TTT CGA GTT TTG CCT CAG CCA ACT AAT CTC GCA ACC CCA AAC ACA AAG CGT TTT AAA

 581  Lys Glu Glu Ile Leu Ser Ser Ser Asp Ile Cys Gln Lys Leu Val Asn Thr Gln Asp Met Ser Ala Ser Gln Val Asp Val Ala
1765  AAA GAA GAA ATT CTT TCC AGT TCT GAC ATT TGT CAA AAG TTA GTA AAT ACT CAG GAC ATG TCA GCC TCT CAG GTT GAT GTA GCT

 609  Val Lys Ile Asn Lys Lys Val Val Pro Leu Asp Phe Ser Met Ser Ser Leu Ala Lys Arg Ile Lys Gln Leu His His Glu Ala
1849  GTG AAA ATT AAT AAG AAA GTT GTG CCC CTG GAC TTT TCT ATG AGT TCT TTA GCT AAA CGA ATA AAG CAG TTA CAT CAT GAA GCA

 637  Gln Gln Ser Glu Gly Glu Gln Asn Tyr Arg Lys Phe Arg Ala Lys Ile Cys Pro Gly Glu Asn Gln Ala Ala Glu Asp Glu Leu
1933  CAG CAA AGT GAA GGG GAA CAG AAT TAC AGG AAG TTT AGG GCA AAG ATT TGT CCT GGA GAA AAT CAA GCA GCC GAA GAT GAA CTA

 665  Arg Lys Glu Ile Ser Lys Thr Met Phe Ala Glu Met Glu Ile Ile Gly Gln Phe Asn Leu Gly Phe Ile Ile Thr Lys Leu Asn
2017  AGA AAA GAG ATA AGT AAA ACG ATG TTT GCA GAA ATG GAA ATC ATT GGT CAG TTT AAC CTG GGA TTT ATA ATA ACC AAA CTG AAT

 693  Glu Asp Ile Phe Ile Val Asp Gln His Ala Thr Asp Glu Lys Tyr Asn Phe Glu Met Leu Gln Gln His Thr Val Leu Gln Gly
2101  GAG GAT ATC TTC ATA GTG GAC CAG CAT GCC ACG GAC GAG AAG TAT AAC TTC GAG ATG CTG CAG CAG CAC ACC GTG CTC CAG GGG

 721  Gln Arg Leu Ile Ala Pro Gln Thr Leu Asn Leu Thr Ala Val Asn Glu Ala Val Leu Ile Glu Asn Leu Glu Ile Phe Arg Lys
2185  CAG AGG CTC ATA GCA CCT CAG ACT CTC AAC TTA ACT GCT GTT AAT GAA GCT GTT CTG ATA GAA AAT CTG GAA ATA TTT AGA AAG

 749  Asn Gly Phe Asp Phe Val Ile Asp Glu Asn Ala Pro Val Thr Glu Arg Ala Lys Leu Ile Ser Leu Pro Thr Ser Lys Asn Trp
2269  AAT GGC TTT GAT TTT GTT ATC GAT GAA AAT GCT CCA GTC ACT GAA AGG GCT AAA CTG ATT TCC TTG CCA ACT AGT AAA AAC TGG

 777  Thr Phe Gly Pro Gln Asp Val Asp Glu Leu Ile Phe Met Leu Ser Asp Ser Pro Gly Val Met Cys Arg Pro Ser Arg Val Lys
2353  ACC TTC GGA CCC CAG GAC GTC GAT GAA CTG ATC TTC ATG CTG AGC GAC AGC CCT GGG GTC ATG TGC CGG CCT TCC CGA GTC AAG

 805  Gln Met Phe Ala Ser Arg Ala Cys Arg Lys Ser Val Met Ile Gly Thr Ala Leu Asn Thr Ser Glu Met Lys Lys Leu Ile Thr
2437  CAG ATG TTT GCC TCC AGA GCC TGC CGG AAG TCG GTG ATG ATT GGG ACT GCT CTT AAC ACA AGC GAG ATG AAG AAA CTG ATC ACC

 833  His Met Gly Glu Met Asp His Pro Trp Asn Cys Pro His Gly Arg Pro Thr Met Arg His Ile Ala Asn Leu Gly Val Ile Ser
2521  CAC ATG GGG GAG ATG GAC CAC CCC TGG AAC TGT CCC CAT GGA AGG CCA ACC ATG AGA CAC ATC GCC AAC CTG GGT GTC ATT TCT

 861  Gln Asn
2605  CAG AAC TGA CCG TAG TCA CTG TAT GGA ATA ATT GGT TTT ATC GCA GAT TTT TAT GTT TTG AAA GAC AGA GTC TTC ACT AAC CTT

2689  TTT TGT TTT AAA ATG AAA CCT GCT ACT TAA AAA AAA TAC ACA TCA CAC CCA TTT AAA AGT GAT CTT GAG AAC CTT TTC AAA CC  ►
```

c

```
yPMS1   MfhhienllietekrckqkeqryipvkylfsmtgIHQINDIDVHRITSGQVITDLTTAVKELVDNSIDANANQIEIIFKD   80
hPMS1   ------------------------------------MKQLPAATVRLLSSSQITSVVSVVKELIENSLDAGATSVDVKLEN   46
hPMS2   Meraesstepaka----------------------LKPIDRKSVHQICSGQVVLSLSTAVKELVENSLDAGATNIDLKLKD   60

yPMS1   YGLESIECSDNGDGIDPSNYEFLALKHYTSKIAKFQDVAKVQTLGFRGEALSSLCGIAKLSVITTTSPPK-ADKLEYDMV   159
hPMS1   YGFDKIEVRDNGEGIKAVDAPVVAMKYYTSKINSHEDLENLTTYGFRGEALGSICCIAEVLITTRTAADNFSTQYVLDGS   126
hPMS2   YGVDLIEVSDNGCGVEEENFEGLTLKHHTSKIQEFADLTQVETFGFRGEALSSLCALSDVTISTCHASAKVGTRLMFDHN   140

yPMS1   GHITSKTTTSRNKGTTVLVSQLFHNLPVRQKEFSKTfkrqftkcltviqgyaiinaaikfsvwnitpkgkknlilstmrn   239
hPMS1   GHILSQKPSHLGQGTTVTALRLFKNLPVRKQFYSTAkkckdeikkiqdllmsfgilkpdlrivfvhnkeviwqksrvsdh   206
hPMS2   GKIIQKTPYPRPRGTTVSVQQLFSTLPVRHKEFQRNikkeyakmvqvlhayciisagirvsctnqlggkrqpvvctggs   220

yPMS1   ssmrkhissvfgaggmrgleevdlvldlnpfknrmlgkytddpdfldldykirvkgyisqnsfgcgrNSKDRQFIYVNKR   319
hPMS1   kmalmsvlgtavmnnmesfqyhseesqiylsgflpkcdadhsftsl----------------------SLPERSFIFINSR   265
hPMS2   psikeniigsvfgqkglqslipfvqlppsdsvceeyglscsdalhnlfyisgfisqcthgvgr-----SSTDRQFFFINRR   295

yPMS1   PVEYSTLLKCCNEVYKTfnnvq-----FFAVFLNLELPMSLIDVNVTPDKRVILLHNERAVIDIFKTTLSDYYNrqelalp   395
hPMS1   PVHQKDILKLIRHHYNLkclkestrlYFVFLLKIDVPTADVDVNLTPDKSQVLLQNKESVLIALENLMTTCYGplpstns   345
hPMS2   PCDPAKVCRLVNEVYHMynrhq----YPFVVLNISVDSECVDINVTPDKRQILLQEEKLLLAVLKTSLIGMFDsdvnkln   371

yPMS1   krmcsqseqqaqkrlktevfdqrstthesdnenyhtarsesnqsnhahfnsttgvidksfgteltsvmdgnytnvtdvig   475
hPMS1   yennktdvsaadivlsktaetqvlfnkvessgknysnvdtsvipfqndmhndesgkntdqclnhqisigdfgyghcssei   425
hPMS2   vsqqplldvegnlikmhaadlekpmvekqdqspslrtgeekkdvsisrlreafslrhttfmkphspktpeprrsplgqkr   451

yPMS1   secevsvdssvvldegnsstptkklpsiktdsqnlsdlnlnnfsnpefqnitspdkarslekvveepvyfdidgekfqek   555
hPMS1   snidkntknafqdismsnvswensqteysktcfissvkhtqsengnkdhidesgeneeeaglensseisadewsrgnilk   505
hPMS2   gmlssstsgaisdkgvlrpqkeavssshgpsdptdraevekdsghgstsvdsegfsipdtgshcsseyaasspgdrgsqe   531

yPMS1   avlsqadglvfvdnechehtndcchqerrgstdteqddsadsiyaeiepveinvrtplknsrksiskdnyrslsdglthr   635
hPMS1   nsvgeniepvkilvpekslpckvsnnnypipeqmnlnesscnkksnvidnksgkvtsaydllsnrvikkpmsasalfvqdh   585
hPMS2   hvdsqekapetddsfsdvdchsnqedtgckfrvlpqptllatpntkrfkkeeilssqsdicqklvntqdmsasqvdvavki   611

yPMS1   kfedeilqynlstknfkeisknqkqmssiiskrkseaqeniiknkdeledfeqgekyltltvskndfkkmevvgqfnlgf   715
hPMS1   rpqflientktsledatlqieelwktlseeeklkyeekatkdlerynsqmkraieqesqmslkdgrkkikptsawnlaqk   665
hPMS2   nkkvvplqfsmsslakrikqlhheaqqsegeqnyrkfrakicpgenqaaedelrkeisktmfaemeiigqfnlgfiitkl   691

yPMS1   iivtrkydnksdlfivdqhaspekynfetlqavtvfksqkliipqpvelsvidelvvldnlpvfekngfklkideeeefg   795
hPMS1   hklktsissnqpkldellqsqiekrrsqnikmvqipfsmknlkinfkkqnkvdleekdepclihnlrfpdawlmtsktevm   745
hPMS2   nedifiydqhatdekynfemlsqhtvlqqqrliapqtlnltaveavlienleifrknqfdfvidenapvteraklislp   771

yPMS1   srvkllslptskqtlfdlgdfnelihlikedgglrrdni----------------------------------------   834
hPMS1   llnpyrveeallfkrllenhklpaepleipimlteslfngshyldvlykmtaddqrysgstylsdprltangfkiklipg   825
hPMS2   tsknwtfgpqdvdelifmlsdspgvmc----------------------------------------------------   798

yPMS1   ----------------------------------------------FCSKIRSMEAMRACRSSIMIGKPLNKKTMTRVVHNLs   871
hPMS1   vsitenyleiegmanclpfygvadlkeilnailnrnakevyecPPRKVISYLEGEAVRLSRQLPMYLSKEDILQDIIYRMk   905
hPMS2   ----------------------------------------------EPSRVKQMEASRACFKSVMIGTALNTSEMKKLITHMg   835

yPMS1   eldkpw--NCPHGRPTMRHLMDLdwssfskdyei                                                  904
hPMS1   hqfgneikECVHGRPFFHHLTYLpetp--------                                                  932
hPMS2   emdhpw--NCPHGRPTMEHIANLgvisqn------                                                  862
```

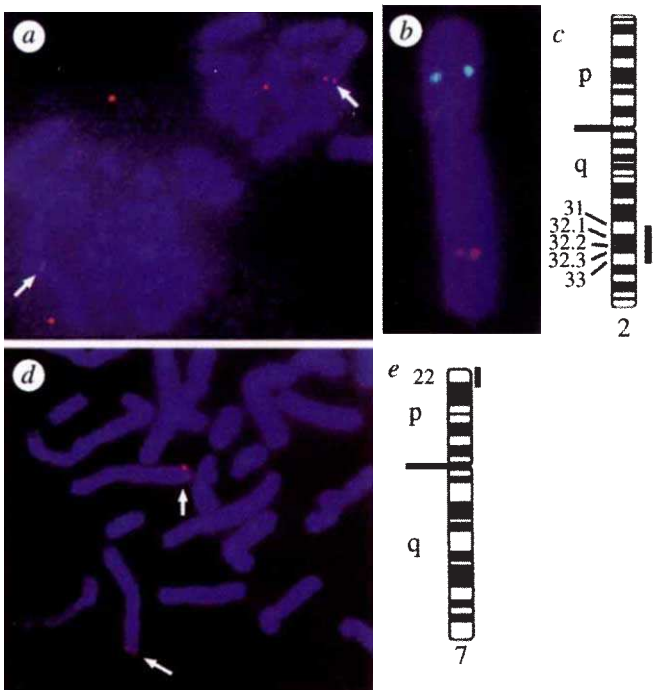

FIG. 2 Chromosomal localization. Metaphase spreads of normal human male chromosomes showing the in situ hybridization of *hPMS1* and *hPMS2*. *a*, Chromosome spread from a single cell showing hybridization of a *hPMS1* genomic clone to the distal q arm of both copies of chromosome 2 (arrows). A doublet signal on chromosome 2q was seen on more than 50 spreads from four separate experiments, while no doublets were detected on any other chromosomes. *b*, Simultaneous hybridization of *hPMS1* (green) and *hMSH2* (red) at chromosome 2p16. *c*, ISCN idiogram showing the predicted position (2q31-33) of *hPMS1* based on fractional length analysis of 14 individual chromosomes[28]. *d*, Chromosome spreads probed with *hPMS2* which hybridizes near the telomere on the p arm of both copies of chromosome 7. A doublet signal on chromosome 7p was seen on more than 40 spreads from three separate experiments, while no doublets were detected on any other chromosomes. *e*, ISCN idiogram of *hPMS2* based on fractional length analysis of 6 individual chromosomes. This analysis indicated that the *hPMS2* gene was located within band 7p22, the most distal band on chromosome 7.

METHODS. A human genomic P1 library (Genome Systems, Inc.) was screened by PCR using primers selected from the cDNA sequence of *hPMS1* and *hPMS2*. Two P1 clones were isolated for *hPMS1* using primers 5′-AAGCTGCTCTGTTAAAAGCG-3′ and 5′-GCACCAGCATCCAAGGAG-3′ which yielded a 133 bp genomic product (nucleotide 58 to 190 of *hPMS1* cDNA). Three P1 clones were isolated for *hPMS2*, using primers 5′-CAACCATGAGACACATCGC-3′ and 5′-AGGTTAGTGAAGACTCTGTC-3′ which yielded a 121 bp genomic product (from nucleotide 2567 to 2687 of *hPMS2* cDNA.) FISH analyses were performed as described[28] except that nick-translated P1 clones were used as probe. Image alignment and fractional length analysis were done using the ISEE software package (Innovision Corp.), Durham, North Carolina.

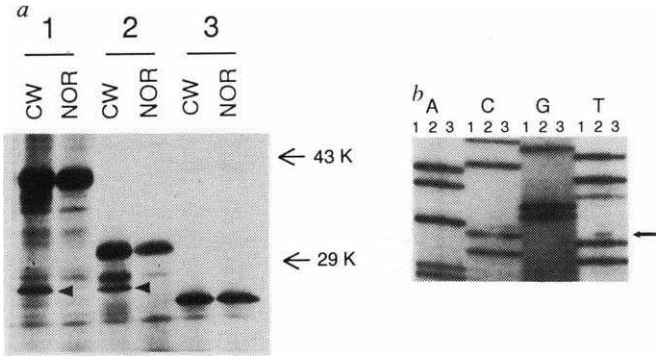

FIG. 3 Mutational Analysis of *hPMS1*. *a*, IVSP mapping of the transcriptional stop mutation in HNPCC patient C.W. Translation of codons 1 to 369 (lanes 1), codons 1 to 290 (lanes 2), and codons 1 to 214 (lanes 3). Lane CW was translated from the cDNA of patient C.W., while NOR was translated from the cDNA of a normal individual. The arrow heads indicate the truncated polypeptide due to the potential stop mutation. The arrows indicate relative molecular mass. *b*, Sequence analysis of C.W. indicates a C to T transition at codon 233 (indicated by the arrow). Lanes 1 and 3 are sequences derived from genomic DNA of control patients; Lane 2 is from genomic DNA of C.W.

METHODS. cDNAs from patient samples were generated from RNA of lymphoblastoid or tumour cells and used as templates in an IVSP assay as described[22]. The *hPMS1* gene was divided into three overlapping segments for the purpose of PCR. The primers for codons 1 to 500 were 5′-GGATCCTAATACGACTCACTATAGGGAGACCACCATGGAACAATTGCCTGCGG-3′ and 5′-CCTGCTCCACTCATCTGC-3′; for codons 270 to 755 were 5′-GGATCCTAATACGACTCACTATAGGGAGACCACCATGGAAGATATCTTAAAGTTAATCCG-3′ and 5′-GGCTTCTTCTACTCTATATGG-3′; and for codons 485 to 933 were 5′-GGATCCTAATACGACTCACTATAGGGAGACCACCATGGCAGGTCTTGAAAACTCTTCG-3′ and 5′-AAAACAAGTCAGTGAATCCTC-3′. The 3′ nested primes used for mapping the stop mutation in patient C.W. were: 5′-AAGCACATCTGTTTCTGCTG-3′ codons 1 to 369; 5′-ACGAGTAGATTCCTTTAGGC-3′ codons 1 to 290; and 5′-CAGAACTGACATGAGAGCC-3′ codons 1 5 to 214. The conditions for all amplifications were 35 cycles at 95 °C for 30 s, 52 °C to 62 °C for 60 to 120 s, and 70 °C for 60 to 120 s, in the buffer described previously[29]. Intron–exon borders of *hPMS1* were determined by cycle-sequencing P1 clones using gamma-^{32}P end labelled primers and SequiTherm polymerase as described by the manufacturer. The primers used to amplify the *hPMS1* exon containing codons 195 to 233 were 5′-TTATTTGGCAGAAAAGCAGAG-3′ and 5′-TTAAAAGACTAACCTCTTGCC-3′, which produced a 215 bp product. The product was directly sequenced using the primer 5′-CTGCTGTTATGAACAATATGG-3′. PCR products were sequenced as described[16].

protein in the absence of any full-length protein (Fig. 4*a*, lane GCx). In order to investigate the nature of this abnormality, the full-length RT-PCR products from G.C.'s normal fibroblasts and tumour xenograft were cloned and sequenced. Analysis of RNA from the normal fibroblasts revealed an altered transcript due to a 1230 bp in-frame deletion of codons 268 to 669 as well as the expected normal transcript. In contrast, analysis of the tumour xenograft identified a second transcript due to a somatic out-of-frame deletion which removed codons 301 to 381. The germline deletion of codons 268 to 669 was also observed in the tumour xenograft, but no normal transcript could be identified. It was likely that these altered transcripts were due to genomic deletions. To test this hypothesis, three sets of primers were used to amplify the *hPMS2* locus, one pair located 5′ to the cDNA deletions (codons 233 to 257), one located 3′ (codons 439 to 472), and one located in the commonly deleted region (codons 347 to 377) (Fig. 4*b*). Both the 5′ and 3′ regions of *hPMS2* could be amplified in all of the samples. The genomic segment containing codons 347 to 377 could be amplified from G.C.'s normal fibroblasts and from control patients but not from the tumour xenograft. These results indicated that the altered transcripts were the result of two intragenic deletions of *hPMS2*, one presumably germline and the other somatic. As expected, the tumour from this patient exhibited microsatellite instability.

During the course of sequence analysis of the *hPMS2* cDNA, we noted a G to A transition in codon 20 which resulted in an arginine to glutamine change. This change was observed in 3 of 18 HNPCC patients and in 5 of 77 controls. We assume this change represents a polymorphism rather than a functional mutation.

The finding that human cells contain at least three *mutL* homologues suggests unique functions for each of these genes. Experimental evidence from *S. cerevisiae* has shown that at least two *mutL* related proteins are required for the major DNA mismatch repair pathway[6,20]. The inactivation of either yeast *mutL* gene (*yMLH1* or *yPMS1*) results in a mutator phenotype[6,20] and the encoded proteins interact in a complex[25]. This view is supported in humans by the observation that inactivation of any one of three *mutL*-related genes can result in HNPCC, presumably due to mismatch repair deficiency. The identification of a

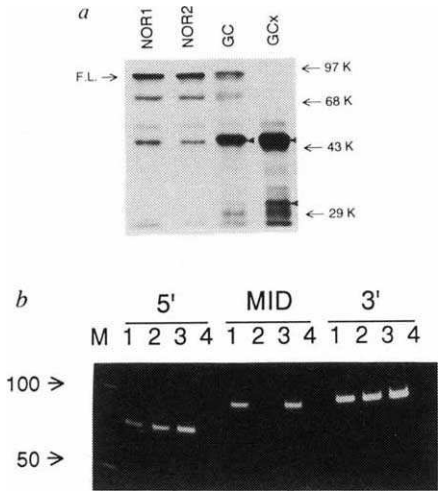

FIG. 4 Mutational analysis of *hPMS2*. *a*, IVSP analysis of *hPMS2* from patient G.C. Lane GC is from fibroblasts of individual G.C.; lane GCx is from the tumour xenograft of patient G.C.; lanes NOR1 and 2 are from normal control individuals. F.L. indicates full-length protein, and the arrowheads indicate the germline truncated polypeptide. A less intense low migrating polypeptide was also seen in all control samples. The arrows indicate relative molecular mass markers. *b*, PCR analysis of patient G.C. shows a deletion of both *hPMS2* alleles in tumour cells. Amplification was done using primers that amplify 5′, 3′, or within (MID) the region deleted in the cDNA. Lane 1, DNA derived from fibroblasts of patient G.C.; lane 2, DNA derived from tumour xenograft of patient G.C.; lane 3, DNA derived from a normal control patient; lane 4, reactions without DNA template. Arrows indicate size in base pairs.

METHODS. IVSP analysis of *hPMS2* was done as described in Fig. 3 legend. In some cases, the PCR products were cloned into the T-tailed cloning vector PCR2000 (Invitrogen) and sequenced with T7 polymerase (United States Biochemical). The *hPMS2* cDNA was amplified as a full-length product or as two overlapping segments. The primers for full-length *hPMS2* were 5′-GGATCCTAATACGACTCACTATAGGGAGACCACCATGGAGCGAGCTGAGAGC-3′ and 5′-AGGTTAGTGAAGACTCTGTC-3′ (codons 1 to 863). For segment 1, the sense primer was the same as above and the antisense primer was 5′-CTGAGGTCTCAGCAGGC-3′ (codons 1 to 472). Segment 2 primers were 5′-GGATCCTAATACGACTCACTATAGGGAGACCACCATGGTGTCCATTTCCAGACTGCG-3′ and 5′-AGGTTAGTGAAGACTCTGTC-3′ (codons 415 to 863). The primers used to analyze the genomic deletion of *hPMS2* in patient G.C. were: For the 5′ region amplification 5′-CAGAAGCAGTTGCAAAGCC-3′ and 5′-AAACCGTACTCTTCACACAC-3′ which produce a 74 bp product containing codons 233 to 257, for the deleted region primers 5′-GAGGAAAAGCTTTTGTTGGC-3′ and 5′-CAGTGGCTGCTGACTGAC-3′ which produce a 93 bp product containing the codons 347 to 377, and for the 3′ region primers 5′-TCCAGAACCAAGAAGGAGC-3′ and 5′-TGAGGTCTCAGCAGGC-3′ which produce a 99 bp product containing the codons 439 to 472 of *hPMS2*.

somatic alteration of the remaining allele in a tumour from a patient with a germline *hPMS2* mutation supports the idea that inactivation of both alleles of a mismatch repair gene is required for tumour formation. □

Received 21 June; accepted 22 July 1994.

1. Lynch, H. T. *et al. Gastroenterology* **104,** 1535–1549 (1993).
2. Peinado, M. A., Malkhosyan, S., Velazquez, A. & Perucho, M. *Proc. natn. Acad. Sci. U.S.A.* **89,** 10065–10069 (1992).
3. Ionov, Y., Peinado, A., Malkhosyan, S., Shibata, D. & Perucho, M. *Nature* **363,** 558–561 (1993).
4. Thibodeau, S. N., Bren, G. & Schaid, D. *Science* **260,** 816–819 (1993).
5. Aaltonen, L. A. *et al. Science* **260,** 812–816 (1993).
6. Strand, M., Prolla, T. A., Liskay, R. M. & Petes, T. D. *Nature* **365,** 274–276 (1993).
7. Parsons, R. P. *et al. Cell* **75,** 1227–1236 (1993).
8. Umar, A. *et al. J. biol. chem.* **269,** 14367–14370 (1994).
9. Peltomaki, P. *et al. Science* **260,** 810–812 (1993).
10. Lindblom, A., Tannergard, P., Werelius, B. & Nordenskjold, M. *Nature Genet.* **5,** 279–282 (1993).
11. Nystrom-Lahti, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **91,** 6054–6058 (1994).
12. Fishel, R. *et al. Cell* **75,** 1027–1038 (1993).
13. Leach, F. S. *et al. Cell* **75,** 1215–1225 (1993).
14. Palombo, F., Hughes, M. & Jiricny, J. *Nature* **367,** 417–418 (1994).
15. Bronner, C. E. *et al. Nature* **368,** 258–261 (1994).
16. Papadopoulos, N. *et al. Science* **263,** 1625–1629 (1994).
17. Adams, M. D. *et al. Science* **252,** 1651–1656 (1991).
18. Frohman, M. A., Dush, M. K. & Martin, G. R. *Proc. natn. Acad. Sci. U.S.A.* **85,** 8998–9002 (1988).
19. Kozak, M. *Cell* **44,** 283–292 (1986).
20. Prolla, T. A., Christie, D. M. & Liskay, R. M. *Molec cell. Biol.* **14,** 407–415 (1994).
21. Liu, B. *et al. Cancer Res.* (in the press).
22. Powell, S. M. *et al. New Engl. J. Med.* **329,** 1982–1987 (1993).
23. Van Der Luijt, R. *et al. Genomics* **20,** 1–4 (1994). (**Initials?**)
24. Dietz, H. C., *et al. Science* **259,** 680–683 (1993).
25. Prolla, T. A. *et al. Science* (in the press).
26. Kramer, W., Kramer, B., Williamson, M. S. & Fogel, S. J. *J. Bact.* **171,** 5339–5346 (1989).
27. Schuler, G. D., Altschul, S. F. & Lipman, D. J. *Proteins Struct. Funct. Genet.* **9,** 180–190 (1991).
28. Johnson, C. V. *et al. Meth. cell. Biol.* **35,** 73–99 (1991).
29. Sidransky, D. *et al. Science* **252,** 706–709 (1991).

ACKNOWLEDGEMENTS. We would like to thank R. Wang, C. Fischer, V. Madike, and the sequencing staff from HGS for technical support. This work was supported by grants from the Clayton Fund, the Academy of Finland, the Finnish Cancer Societies, the Sigrid Juselius Foundation, the Folkhalsan Institute of Genetics, the Cooperative Human Tissue Network (NCI) and the NIH. N.C.N. and N.P. contributed equally to this work.

Crystal structure of an isoleucine-zipper trimer

Pehr B. Harbury*†, Peter S. Kim† & Tom Alber‡

* Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA
† Howard Hughes Medical Institute, Whitehead Institute, Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA
‡ Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA

SUBUNIT oligomerization in many proteins is mediated by short coiled-coil motifs[1,2]. These motifs share a characteristic seven-amino-acid repeat containing hydrophobic residues at the first (a) and fourth (d) positions. Despite this common pattern, different sequences form two-, three- and four-stranded helical ropes. We have investigated the basis for oligomer choice by characterizing variants[3] of the GCN4 leucine-zipper dimerization domain that adopt trimeric or tetrameric structures in response to mutations at the a and d positions. We now report the high-resolution X-ray crystal structure of an isoleucine-containing mutant that folds into a parallel three-stranded, α-helical coiled coil. In contrast to the dimer and tetramer structures[3,4], the interior packing of the trimer can accommodate β-branched residues in the most preferred rotamer at both hydrophobic positions. Compatibility of the shape of the core amino acids with the distinct packing spaces in the two-, three- and four-stranded conformations appears to determine the oligomerization state of the GCN4 leucine-zipper variants.

The GCN4-pII peptide, which differs from the wild-type GCN4 leucine zipper by isoleucine substitutions at four **a** and four **d** positions (Fig. 1*a*; Table 1), forms a parallel trimer in solution[3]. A similar isoleucine repeat occurs in the trimeric σ1 protein of reovirus serotype 1 (refs 5, 6) in haemagglutinins from several strains of human parainfluenza virus[7,8] and in trout axonemal dynein[9]. The trimeric coiled coils from influenza haemagglutinin[10,11] and the yeast heat-shock factor[12] (HSF) resemble the GCN4-pII sequence at the **d** positions of the heptad repeat, which contain predominantly β-branched residues (IIIIL in the five heptads of HSF).

The X-ray crystal structure of GCN4-pII at 1.8 Å resolution (Fig. 1) shows that three α-helical peptide monomers wrap in a gradual left-handed superhelix. The superhelix describes a cylinder that is ~24 Å wide and ~48 Å long. The isoleucine residues point into the core of the trimer, and cross-sectional layers containing isoleucine at the **a** positions alternate with layers containing isoleucine at the **d** positions (Fig. 1*d*). In seven of eight layers, the dihedral angles χ_1 and χ_2 of the isoleucine side chains are approximately −60,180, corresponding to the most abundant rotamer[13]. The isoleucine residues in the C-terminal **a** layer assume dihedral angles near +60,180. Residues at positions **e** and **g** pack against the isoleucines at **d** and **a**, respectively, to complete the hydrophobic core. Interhelical salt bridges form with high frequency between charged side chains at the **g** position of one heptad and the **e** position of a succeeding heptad (for

TABLE 1 Helix–helix interactions*

	Dimer		Trimer	Tetramer	Globular proteins[17]
GCN4 peptide variant†	p1‡ and pI	L	pI I	pL I	
Residues at four **a** positions	V(N)	I	I	L	
Residues at four **d** positions	L	L	I	I	
Superhelix parameter					
Supercoil radius, R_0(Å)	4.9		6.7	7.6	
Residues per supercoil turn, ω_0	100		118	139	
Supercoil pitch (Å)	148		175	205	
Radius of curvature (Å)	118		124	149	
Superhelix crossing angle (χ)	23.4°		26.8°	26.0°	
Position **a** orientation angle, ϕ	21.6°		20.4°	19.8°	
α-Helix parameter[17]					
Residues per α-helix turn, n	3.62		3.60	3.59	3.64 ± 0.18
Rise/residue, d (Å)	1.51		1.53	1.52	1.51 ± 0.12
α-Helix radius (Cα), R_1(Å)	2.28		2.24	2.26	—
Pairwise helix-crossing angle, Ω	23.4°		23.2°	18.3°	19° ± 24°
Pairwise interhelix distance, D (Å)	9.8		11.5	10.6	10.2 ± 2.0
Interhelical salt bridges					
g to **e**	3 of 6		8 of 9	5 of 12	
g to **b** and **c** to **e**	0 of 8		1 of 12	9 of 16	

* Superhelical characteristics were obtained by fitting the Cα backbones to a supercoil parametrization suggested by Crick[22] (Fig. 1*a*). The radius (R_0), frequency (ω_0) and pitch of the superhelix, and the radius (R_1) and phase (ϕ) of the α-helix were treated as variables. The frequency of the α-helix was fixed at 4π radians per 7 amino acids to preserve the periodicity of the heptad structure. α-helix parameters were calculated as defined in ref. 17. The properties of α-helices in globular proteins were taken from that reference and are expressed as the mean value ±2 standard deviations. The values of D and Ω for globular proteins are averages for the 3–4 ridges-into-grooves packing class[17]. Consistent with earlier analyses[3,22,23], the superhelical radius and pitch, and the radius of helix curvature increase with the number of strands in the coiled coil.

† The leucine zipper variants were derived from the wild-type GCN4 coiled-coil peptide, GCN4-p1 (ref. 24), by collective mutation to the indicated residues at four **a** positions (dashed box in Fig. 1*a*) and four **d** positions (dashed oval in Fig. 1*a*)[3].

‡ Geometric properties of the dimer conformation were obtained from the structure of GCN4-p1 (ref. 4) because the crystallographic analysis of GCN4-pIL has not been completed. The GCN4-p1 sequence contains four leucines at the varied **d** positions, and three valines and one asparagine at the varied **a** positions. Interestingly, replacement of the asparagine residue (Asn 16) by valine generates a peptide, GCN4-pVL, that forms a mixture of dimers and trimers[3]. This result indicates that the buried polar side chain of Asn 16 directs dimer formation.