Alternative Splicing of MLH1 Messenger RNA in Human Normal Cells

François Charbonnier, Cosette Martin, Michel Scotte, Louis Sibert, Viviane Moreau, and Thierry Frebourg

Laboratoire de Génétique Moléculaire [F. C., C. M., V. M., T. F.], Service de Chirurgie Générale et Digestive [M. S.], and Service d’Urologie [L. S.]. Centre Hospitalo-Universitaire de Rouen, 1 rue de Germont, 76031 Rouen Cédex, France

Abstract

The hMLH1 protein, composed of 756 amino acids, is the human homologue of the bacterial DNA mismatch repair protein MutL, and germ line mutations of the hMLH1 gene have been identified in kindreds with hereditary nonpolyposis colorectal cancer. We have detected three alternatively spliced forms of hMLH1 mRNA in normal lymphocytes and tissues. One of the spliced forms lacks the coding region of hMLH1 from codons 227 to 295 and the two other transcripts are predicted to encode two truncated proteins retaining the 264 and 226 N-terminal amino acids of hMLH1, respectively. The biological significance of this alternative splicing remains to be established.

Introduction

The HNPCC or Lynch syndrome is one of the main causes of familial colorectal cancer (1). This autosomal dominant syndrome represents a genetic predisposition not only for colorectal cancers but also for other cancers such as adenocarcinomas of the endometrium, stomach, ovary, small bowel, and hepatobiliary tract and transitional cell carcinomas of the urinary tract (1). Recent studies (2–6) have established that the genes involved in HNPCC are the human homologues of the prokaryotic MutS and MutL genes which are essential in Escherichia coli for the initiation of the repair of mismatched DNA (7, 8). In E. coli, the first steps of the methyl-directed repair of mismatch made during DNA replication and genetic recombination are (a) the binding of the MutS protein to the mismatch, (b) the addition of the MutL protein to the MutS heteroduplex which activates the endonuclease activity of the MutH protein, and (c) the incision at a hemimethylated (GATC) sequence of the transient unmethylated strand (7, 8). The human homologue of the MutS gene is the hMSH2 gene which is located on chromosome 2p22–21 (2, 3). Three human homologues of the MutL gene, hMLH1, hMLH2, hMPS1, and hPMS2 (6), respectively, located on chromosomes 3p21, 2q31–33, and 7p22, have been cloned. Germ line mutations of the hMSH2, hMLH1, hMPS1, and hPMS2 genes have been reported in HNPCC pedigrees and the involvement of these mutations in HNPCC has been established by the cosegregation of the mutant allele with cancer (2–6, 9–11, 13). Linkage studies have suggested that most of the HNPCC cases result from germ line alterations of the hMSH2 and hMLH1 genes (5). The hMLH1 gene is composed of 19 exons (12, 13) and the hMLH1 cDNA is predicted to encode a protein of 756 amino acids (4, 5). We report here that the hMLH1 mRNA is alternatively spliced and that these alternatively spliced forms potentially encode shorter hMLH1 proteins.

Materials and Methods

Amplification of hMLH1 cDNAs. Peripheral blood lymphocytes purified from 10 ml of healthy blood donors and macroscopically normal tissues collected after surgical removal or endoscopy were placed directly in 400 µl RNA lysis buffer (Pharmacia Biotech, Uppsala, Sweden) and stored at −20°C. mRNA was extracted using the QuickPrep Micro mRNA purification kit (Pharmacia) and was resuspended into 400 µl water. Random hexamer-primed cDNAs were synthesized from 20 µl of the cDNA reaction using as sense primer, MA2F (CAG GAA ACA GCT ATG ACC TTC ACT ACÁ CAÁ TOC AGG) or MAF (CGG GAT CCC GGC ATC TAG ACG TTT CCT TGG C), and as antisense primer MAR1 (CGG AAT TCC GCA TCA AGC TTC TGT TCC CG; Fig. 1). As denoted by the underlining, the primers contained either an additional M13 reverse sequence (MA2F), a BamHI restriction site (MAF), or an EcoRI restriction site (MAR1). PCR was performed in a final volume of 100 µl containing 0.5 µM of primer and 2.5 units of Taq DNA polymerase from Perkin Elmer/Cetus, Foster City, CA) and an Applied Biosystems model 373A automated sequencer.

Cloning and Sequencing of hMLH1 cDNAs. cDNAs amplified with the MA2F and MAR1 primers were digested with SspI [an SspI restriction site is located at nucleotide 684 of the hMLH1 cDNA (4)] and EcoRI restriction enzymes and cDNAs amplified with the MAF and MAR1 primers were digested with BamHI and EcoRI restriction enzymes. Fragments were then purified by electrophoresis on low-melt agarose and ligated into the plBlue-script SK(+) vector (Stratagene) cut with the corresponding restriction enzymes. After bacterial transformation, individual clones were sequenced on both strands using the PRISM Ready Reaction Dye Primer sequencing kit (Applied Biosystems, Perkin Elmer/Cetus, Foster City, CA) and an Applied Biosystems model 373A automated sequencer.

Measurement of the Relative Expression of the Different hMLH1 Spliced Forms. The relative amounts of the different transcripts expressed in lymphocytes were estimated by RT-PCR using the primer MAR1 and a MA2F primer end labeled with the fluorescein dye C6-FAM (6-carboxyfluorescein) from Applied Biosystems. The PCR product (0.25 µl) and 0.5 µl of internal lane size standard (Applied Biosystems GeneScan-2500 Rox) were loaded on a 6-cm 6% (w/v) denaturing polyacrylamide gel. The electrophoresis was performed at 600 V for 5 h and data were analyzed using the Gene Scanner Model 672 Fluorescent Fragment Analyzer (Applied Biosystems).

Results

The expression of the hMLH1 gene in normal lymphocytes and tissues was analyzed by RT-PCR using the MA2F and MAR1 primers which allow an amplification of codons 190–388 of the hMLH1 cDNA (Fig. 1). This analysis revealed, beside the expected band at 663 base pairs, several bands of lower molecular weight (Fig. 2). Cloning and sequencing of these cDNAs derived from the lymphocytes of one normal subject revealed four different partial cDNAs (that we named I, II, III, and IV) which presumably result from an alternative splicing of the hMLH1 mRNA between codons 226 and 347 (Fig. 1). The cDNA I had a sequence identical to the one previously reported (4, 5). The cDNA II contained a 207-base pair deletion of the coding region from the third nucleotide of codon 226 to the second nucleotide of codon 295 and therefore lacked exons 9
and 10 of the hMLHI gene (12). The cDNA III had a 248-base pair deletion from the second nucleotide of codon 264 to the third nucleotide of codon 346 and lacked exons 10 and 11. The cDNA IV had a 361-base pair deletion corresponding to exons 9, 10, and 11. RT-PCR performed on lymphocytes using the primers MAF and MARI, which allowed an amplification of codons 1–388 of hMLHI (Fig. 1), revealed, as expected, several bands (data not shown). Cloning and sequencing of the PCR products derived from the lymphocytes of one normal subject indicated that the sequences of the cDNAs I, II, III, and IV were identical between codons 1 and 226. The deletion observed in transcript II preserves the open reading frame, and translation of this transcript would result in a MLHI protein lacking 69 amino acids between codons 227 and 295. The deletion observed in transcript III produces a frameshift with a new stop codon located 42 nucleotides downstream and translation of this transcript would result in a truncated MLHI protein of 277 amino acids containing 13 new amino acids (Fig. 1). In transcript IV, the deletion also produces a frameshift with a new stop codon located 59 nucleotides downstream and translation of this transcript would result in a truncated MLHI protein of 245 amino acids with 19 new amino acids (Fig. 1). Transcripts III and IV were clearly observed not only in lymphocytes but also in different tissues including colon, stomach, breast, bladder, and skin (Fig. 2). RT-PCR analysis with a fluorescent primer showed that transcript I was the most abundant form in lymphocytes and revealed marked differences in the relative abundances of the four hMLHI transcripts among normal individuals (Fig. 3 and Table 1).

Discussion

RT-PCR analysis allowed us to show that the hMLHI mRNA is alternatively spliced in normal cells and potentially encodes several MLHI isoforms of different sizes. The alternative spliced forms of hMLHI mRNA that we have described may have no biological importance. Nevertheless, the expression level of the different transcripts measured in lymphocytes (Table 1) and the detection of these isoforms may have clinical significance.
alternatively spliced forms in several normal tissues (Fig. 2) may indicate a possible role of this splicing in MLH1 function. The study of the biological consequence of this splicing will require knowledge of the different functional domains of the hMLH1 protein. In E. coli, the MutL protein has been shown to interact with the MutS hetero-duplex-DNA complex in the presence of ATP, but the precise biochemical function of this protein is at the present time unknown (7, 8, 14). Two of the transcripts of the hMLH1 gene (III and IV), which were clearly detectable in different tissues (Fig. 2), are predicted to encode truncated proteins retaining only the 264 and 226 N-terminal amino acids of the hMLH1 protein, respectively (Fig. 1). The hMLH1 protein exhibit a high degree of homology with the E. coli MutL protein and with the products of the MutL-related genes identified in Salmonella typhimurium, Streptococcus pneumoniae, and Saccharomyces cerevisiae (4, 5). The greatest homologies between the human and yeast MLH1 proteins are within the N-terminal region, from amino acids 8–229 (57% of identity) and within the COOH terminus, from amino acids 739–756 (72% of identity). The conserved region at the NH2 terminus of the hMLH1 protein contains domains which are homologous to the hPMS1 and hPMS2 proteins (6). The observation that the 226 N-terminal amino acids of the protein are highly conserved and retained in the predicted products of the different hMLH1 alternatively spliced forms might suggest that this region corresponds to a functionally important domain of the MLH1 protein. The alternative splicing of the hMLH1 gene in lymphocytes may also have an important practical implication. The majority of the germ line alterations of the mismatch repair genes reported so far in HNPCC families are predicted to result in truncated proteins (2–6, 9–11). These alterations can efficiently be detected by the in vitro transcription/translation of the PCR-amplified cDNAs and analysis of the size of the resulting polypeptides (5, 6, 10, 12). In some cases the alternative splicing of the hMLH1 gene that we have documented in peripheral blood lymphocytes may complicate the detection of germ line hMLH1 mutations based on this strategy.

Acknowledgments

We thank Drs. Pascal Joly, Jean-Pierre Julien, and Luc Thiberville for their assistance with samples.

References

Alternative Splicing of \textit{MLH1} Messenger RNA in Human Normal Cells

Françoise Charbonnier, Cosette Martin, Michel Scotte, et al.