Defective Mismatch Repair, Microsatellite Mutation Bias, and Variability in Clinical Cancer Phenotypes

Sandeep N. Shah¹,², Suzanne E. Hile¹, and Kristin A. Eckert¹,²

Abstract
Microsatellite instability is associated with 10% to 15% of colorectal, endometrial, ovarian, and gastric cancers, and has long been used as a diagnostic tool for hereditary nonpolyposis colorectal carcinoma–related cancers. Tumor-specific length alterations within microsatellites are generally accepted to be a consequence of strand slippage events during DNA replication, which are uncorrected due to a defective postreplication mismatch repair (MMR) system. Mutations arising within microsatellites associated with critical target genes are believed to play a causative role in the evolution of MMR-defective tumors. In this review, we summarize current evidence of mutational biases within microsatellites arising as a consequence of intrinsic DNA sequence effects as well as variation in MMR efficiency. Microsatellite mutational biases are generally not considered during clinical testing; however, we suggest that such biases may be clinically significant as a factor contributing to phenotypic variation among microsatellite instability–positive tumors.

Introduction
Microsatellites are short, repetitive elements in genomic DNA consisting of one to six bases per repeat unit, and are nonrandomly distributed throughout the human genome (for reviews, see refs. 1, 2). Microsatellites located in intragenic regions (promoters, 3′-untranslated regions, and introns) can be important regulators of gene expression by influencing transcription rate, RNA stability, splicing efficiency, and RNA-protein interactions. Because microsatellite sequences are highly polymorphic in human populations, genetic changes in some microsatellites can be important evolutionarily by affecting phenotype. Microsatellites located within intergenic regions might also have functional roles in chromatin organization and recombination.

Postreplication mismatch repair (MMR) is a critical mechanism for maintaining microsatellite stability through the correction of base substitution mismatches and insertion/deletion events (for review, see ref. 3). Premutational intermediates are detected and processed by heterodimers of the MutS and MutL family of proteins. The MutSo complex (MSH2 and MSH6) can recognize base mispairs as well as single-base insertion/deletion loops (IDL) in either the parental or nascent strand. MutSβ (MSH2 and MSH3) can recognize IDLs of two to four bases in addition to single-base IDLs. MutL heterodimers are recruited by MutS complexes to the mismatch or IDL, and consist of MLH1 dimerized with either PMS2 (MutLα), PMS1 (MutLβ), or MLH3 (MutLγ). MutLα is regarded as the major MutL complex for the repair of both large and small IDLs, whereas MutLγ and MutLβ have little or no known role in MMR, respectively.

Clinical Phenotype of Microsatellite Mutation
Lynch syndrome, also known as hereditary nonpolyposis colorectal cancer (HNPPC), is characterized by early onset tumors in the proximal colon as well as extracolonic regions (for review, see ref. 4). Lynch syndrome results from the germ line inheritance of mutations in MMR genes and subsequent MMR deficiency arising during neoplastic cell evolution. Such MMR deficiency can be readily diagnosed as tumor-specific microsatellite instability (MSI), using a panel of defined microsatellite markers. Patients that are homozygotes or compound heterozygotes for mutations in MSH2, MSH6, MLH1, or PMS2 genes could present with childhood hematologic and brain malignancies as well as features of neurofibromatosis (e.g., café au lait spots), in addition to early onset gastrointestinal neoplasms.

Germline mutations in patients with a history of HNPPC are most frequently seen in the MSH2 or MLH1 genes. Due to partial redundancy and differences in substrate specificity, MSH6 (MutSo specific) alterations are less common in HNPPC, whereas genetic predisposition to HNPPC has not been shown for MSH3 (MutSβ specific) mutations (5). Similarly, PMS2 (MutLα specific) mutations are rarely observed in patients with HNPPC, as compared with MLH1 mutations, even though PMS2 is the primary binding partner of MLH1. This discrepancy in the frequency of MLH1 versus PMS2 mutations might be due, in part, to the redundancy between PMS2 and MLH3 activities because MLH3 has similar biochemical properties.

References

©2010 American Association for Cancer Research.
to PMS2 (6). However, investigations of MLH3 (MutLβ specific) and PMS1 (MutLβ specific) have shown only a minor role in microsatellite stability (3), and it is unclear whether mutations in these genes are involved in HNPPC-like phenotypes (5).

Although MMR proteins act in a single biochemical pathway, loss of individual components results in both similar and distinct clinical phenotypes. Patients carrying heterozygous mutations in either the MLH1 or MSH2 genes display classic HNPPC, whereas patients carrying homozygous MSH6 and PMS2 mutations usually display a decreased penetration and later onset of HNPPC (5). Genotypic-phenotypic correlations for HNPPC are complicated, as patients carrying different variant alleles within the same gene could have distinct clinical phenotypes. For example, a recent study described a patient with an MLH1-G67E mutation and family history of HNPPC that displayed both classic and atypical tumors, distinct from families with MLH1-G67R mutations that only show classic tumors (7). Clinical phenotypes could also vary widely among patients having the same allelic mutation. For example, many carriers with an MSH2-A636P allele displayed only colorectal carcinomas, whereas other patients also developed endometrial or ovarian carcinomas (8). Variable presentations of tumor types among patients with HNPPC might reflect the stochastic nature of genetic alterations (mutation timing, frequency, and specificity) that result from the mutator phenotype initiated by defective MMR.

**Microsatellite Mutational Biases Due To Intrinsic DNA Features**

Mutation rates of common microsatellites found in the human genome are extremely variable. The combined mutation rate estimates from analyses of parent-child allele transmissions and experimental determinations in somatic, nontumorigenic human cells range from a low of $10^{-6}$ to a high of $10^{-2}$ mutations per locus per generation (for reviews, see refs. 1, 9). Multiple extrinsic factors influence microsatellite mutation rates, including flanking sequence, recombination rate, sex, and age. Undoubtedly, however, the vast majority of the observed variation in mutagenesis is due to features intrinsic to the repeated DNA sequence itself, including motif unit size (mononucleotide, dinucleotide, trinucleotide, etc.), length (number of units), and sequence composition (1, 9, 10). In general, motif size is inversely related to microsatellite allele mutation frequency, whereas motif length is directly related to microsatellite mutation frequency. Motif sequence can induce mutational variation in several ways. First, the precise sequence composition of an allele affects thermostability and DNA secondary structure potential, both of which may be factors contributing to the significant differences in microsatellite allele mutation rates observed in human cells (11). Second, the purity of a microsatellite directly affects mutability, as interrupted alleles have a significantly lower mutation rate. Third, the complexity of a microsatellite locus (simple, compound, or complex) has been implicated as a factor in mutability (1). In addition to differences in absolute mutation frequency, the types of mutational events arising within microsatellite alleles are dependent on intrinsic DNA features. Mathematical modeling using human genome data has shown that expansions are most prevalent in short microsatellites, whereas long microsatellites tend to contract (1). Nontumorigenic human cells also display distinct mutational biases for contractions versus expansion mutations within dinucleotide and tetranucleotide microsatellites due to differences in motif sequence composition (9, 11).

Mechanistically, microsatellite alleles expand or contract through strand slippage events during DNA synthesis. One potential source of the above mutational biases may be differential stabilization of slipped strand intermediates through formation of stable secondary DNA structures within microsatellite alleles. In this mechanism, both the DNA sequence content and the DNA strand location (leading versus lagging strand) of a microsatellite contribute to mutational bias. DNA repair has also been shown to impart mutation biases.

**Microsatellite Mutational Biases Due To Defective MMR**

The repair specificity of individual MMR proteins is an especially important consideration for disease phenotype. DNA repair by each MutS heterodimer is substrate-specific within mononucleotide, dinucleotide, or tetranucleotide repeats: MutSα is expected to repair premutational intermediates only within mononucleotide repeats, whereas MutSβ potentially repairs premutational intermediates within mononucleotide, dinucleotide, trinucleotide, and tetranucleotide repeats (12). In yeast and *Escherichia coli* models, the efficiency of MMR within microsatellites is dependent on motif size, length, and sequence composition of an allele (9). Mutation rate variations among GT/CA loci within the yeast genome have been attributed to differential MMR efficiencies (13). We have recently shown that differential MMR efficiency by PMS2 contributes to mutation rate biases at mononucleotide, dinucleotide, and tetranucleotide alleles in nontumorigenic human cells (14). Thus, available experimental data show that MMR contributes to mutation frequency biases within microsatellites.

We and others also have observed differences in the mutational specificity of microsatellites due to defects in specific MMR genes (Table 1). In mice, a deficiency of MSH2 gives rise to 78% 1-bp deletions in a mononucleotide allele; in contrast, loss of MSH3 or MSH6 results in only 61% or 47% deletion mutations within mononucleotides, respectively (15). For dinucleotide GT/CA alleles, loss of yeast MSH3 results in a bias favoring deletions, whereas loss of yeast MSH6 results in the opposite bias, favoring insertions. In MSH2-deficient yeast, the majority of mutations at [AG/TC]$_n$ within the *APC* gene are insertions, whereas deletions are more frequent in an [A/T]$_n$ tract of the same gene (16). Human *APC* in MMR-defective colorectal tumors displays a reverse trend in which deletions are readily observed within the [AG/TC]$_n$ tract, whereas insertions are common at [A/T]$_n$ (16). The human *APC* data was obtained from a mutation database.
representing patients of various MMR backgrounds, and as such, mutation spectra are a combination of various MMR defects. The discrepancy between yeast and human data could also indicate selection for the tumor phenotype.

Loss of specific members of the MutL family also produces distinct mutational biases. For mononucleotide [G/C]7–8 alleles in mice, 80% of mutations in the absence of MLH1 were deletions, compared with only 64% with PMS2 loss (Table 1; ref. 15). The difference between MutL members (MLH1 and PMS2) was even more dramatic for an AC/TG dinucleotide locus, in that MLH1-deficient mice displayed a bias in favor of deletions whereas PMS2-deficient mice displayed a bias in favor of expansions (17). In our recent publication, we showed that a bias in favor of expansions within microsatellites was also observed in PMS2-deficient human cells (14). As shown in Table 1, deficiency of PMS2 increases expansion mutations as compared with deletions within the [TTCC/AAGG]9 and [TTTC/AAAG]9 tetranucleotide repeats, as well as within a [GT/CA]10 dinucleotide repeat. This bias may be significant in vivo, as loss of PMS2 allows nuclear localization of MLH3 (6), thereby potentially changing the mutation profile. We proposed that repair proteins such as MLH1-MLH3 might be able to compensate for PMS2 loss, as MMR components have very distinct effects during DNA repair. Consistent with this interpretation, knockout mice with PMS2/MLH3 deficiency display much greater MSI than PMS2-deficient mice (18). Taken together, these experimental data show that loss of PMS2 generates microsatellite mutational biases that are quite distinct from those generated by loss of MLH1.

### Target Genes of MSI

Three mechanisms have been described by which alterations in microsatellite allele length directly affect neoplastic progression: (a) gene inactivation due to frameshift mutations within exonic microsatellites; (b) gene activation due to enhanced transcription factor binding of microsatellites within promoter regions; and (c) microsatellite allele length differences that affect gene expression levels.

Mutation of mononucleotide microsatellites within exonic regions has been extensively analyzed in MSI colorectal tumors (reviewed in ref. 19). Classes of genes known to be altered in MSI tumors include DNA repair, cell growth and signaling, apoptosis, and transcription factors. For example, the TGFBR1 gene contains an (A)10 repeat which is mutated in 86% of MSI-positive tumors, compared with just 0.6% of MSI-negative tumors (20). It is also interesting to note that several of the MMR genes, i.e., MSH6, MSH3, PMS2, and MLH3, also contain mononucleotide microsatellites. This property has been proposed to allow fine-tuning of mutation rates for adaptation during evolution (21) and may also play a role in the immune response (22). MMR is highly involved in class switch recombination and somatic hypermutation, and thus, control of immunoglobulin diversity. Switch regions contain highly repetitive G/C-rich sequences, and loss of individual MMR proteins has been shown to differentially regulate the frequency of recombination within these regions.

Intragenic microsatellites located within introns or 5′- and 3′-untranslated regions should also be considered as potential target genes for MSI. For example, experiments to elucidate the mechanism of gene regulation by the EWS/FLI oncogenic transcription factor found that a minimum of four consecutive GGAA sequences need to be present in the promoter of target genes for transcriptional activity. Importantly, transactivation increased with each additional unit of GGAA, and differences in microsatellite size and sequence have been suggested to be a factor determining susceptibility to Ewing’s sarcoma (23). Microsatellite allele length polymorphisms are known to be correlated with differential gene

---

**Table 1. Microsatellite mutational biases due to loss of specific MMR gene functions**

<table>
<thead>
<tr>
<th>MMR gene deficiency</th>
<th>Species</th>
<th>STR</th>
<th>Del/Ins</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSH2</td>
<td>Mouse</td>
<td>(G/C)7–8</td>
<td>4:1</td>
<td>(15)</td>
</tr>
<tr>
<td></td>
<td>Yeast</td>
<td>(GT/CA)16.5</td>
<td>1:1</td>
<td>(12)</td>
</tr>
<tr>
<td>MSH6</td>
<td>Mouse</td>
<td>(G/C)7–8</td>
<td>1:1</td>
<td>(15)</td>
</tr>
<tr>
<td></td>
<td>Yeast</td>
<td>(G/C)18</td>
<td>3:1</td>
<td>(12)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(GT/CA)16.5</td>
<td>1:3.5</td>
<td>(12)</td>
</tr>
<tr>
<td>MSH3</td>
<td>Mouse</td>
<td>(G/C)7–8</td>
<td>2:1</td>
<td>(15)</td>
</tr>
<tr>
<td></td>
<td>Yeast</td>
<td>(GT/CA)16.5</td>
<td>4:1</td>
<td>(12)</td>
</tr>
<tr>
<td>MLH1</td>
<td>Mouse</td>
<td>(G/C)7–8</td>
<td>4:1</td>
<td>(15)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(A/T)23–27</td>
<td>18:1</td>
<td>(17)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(AC/TT)22–33</td>
<td>2:1</td>
<td>(17)</td>
</tr>
<tr>
<td>PMS2</td>
<td>Mouse</td>
<td>(G/C)7–8</td>
<td>2:1</td>
<td>(15)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(A/T)23–27</td>
<td>6:1</td>
<td>(17)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(AC/TT)22–33</td>
<td>1:2</td>
<td>(17)</td>
</tr>
<tr>
<td>Human</td>
<td></td>
<td>(GT/CA)10</td>
<td>1:11</td>
<td>Unpublished observations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(TTCC/AAGG)9</td>
<td>1:21</td>
<td>(14)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(TTTC/AAAG)9</td>
<td>1:4</td>
<td>(14)</td>
</tr>
</tbody>
</table>
expression. A relevant example for neoplasia is the epidermal growth factor receptor gene, which contains a polymorphic (CA)_{14,21} unit repeat within intron 1. Variation in the number of (CA) units causes differing levels of epidermal growth factor receptor gene expression and concomitant protein production (shorter alleles produce greater protein), which may be a predictor of clinical outcome in breast cancer (24).

The mutator phenotype model of cancer, which is appropriate for the description of MMR-deficient tumors, postulates that tumors evolve as a heterogeneous collection of genetically distinct cells, each having different, but frequently overlapping, patterns of mutation (25). Direct analyses of mononucleotide alleles in MSI colorectal carcinomas have shown that strand slippage mutations are as abundant in short, intergenic microsatellites as within intragenic, coding repeats (26). Thus, in addition to microsatellite sequence changes that are phenotypically selected during tumor cell evolution, a high frequency of random ("passenger") microsatellite changes is expected to arise throughout the genome of individual MMR-deficient tumor cells.

Perspective

Long regarded as "junk DNA," microsatellites are emerging as important regulators of genome function and gene expression. Historically, the primary clinical use of microsatellites has been as markers of genome instability. However, a more direct function of microsatellite instability is to create genetic diversity for tumor cell evolution or immune system adaptability. Mutational biases (e.g., high versus low mutability, insertion versus deletion events) within microsatellites are caused by intrinsic DNA features and by loss of a specific MMR protein. Such biases will directly determine the complement of microsatellite-associated target genes that are affected in MSI tumors. In addition, individual germ line polymorphisms in microsatellite sequence purity and length will affect the inherent mutability of a particular microsatellite allele. We postulate that somatic mutational biases, together with germ line microsatellite polymorphisms, may underlie the tissue-specific and individual-specific pattern of tumor formation observed among patients with HNPCC and other MSI. The unique pattern of mutational events—error type, affected unit size, and length of the microsatellite—arising within an MMR-defective tumor cell population will thus shape an individual's tumor genome landscape, leading to a distinct clinical phenotype. Understanding the genotypic-phenotypic relationships in MMR-deficient tumors is necessary to develop better predictive models of disease progression and novel therapeutic approaches.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Grant Support

NIH grant no. RO1 CA100060 (K. Eckert) and the Jake Gittlen Cancer Research Foundation.

Received 8/14/09; revised 9/22/09; accepted 9/24/09; published OnlineFirst 1/12/10.

References

Defective Mismatch Repair, Microsatellite Mutation Bias, and Variability in Clinical Cancer Phenotypes

Sandeep N. Shah, Suzanne E. Hile and Kristin A. Eckert

Cancer Res  Published OnlineFirst January 12, 2010.

Updated version  Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-09-3049

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/early/2010/01/12/0008-5472.CAN-09-3049. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.