Both microsatellite length and sequence context
determine frameshift mutation rates in defective
DNA mismatch repair

Heekyung Chung1, Claudia G. Lopez1, Joy Holmstrom1, Dennis J. Young2, Jenny F. Lai1,
Deena Ream-Robinson1 and John M. Carethers1,2,3,*

1Department of Medicine and 2Rebecca and John Moores Comprehensive Cancer Center, University of California,
San Diego, CA, USA and 3Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA

Received December 22, 2009; Revised and Accepted April 13, 2010

It is generally accepted that longer microsatellites mutate more frequently in defective DNA mismatch repair
(MMR) than shorter microsatellites. Indeed, we have previously observed that the A10 microsatellite of trans-
forming growth factor beta type II receptor (TGFBR2) frameshifts −1 bp at a faster rate than the A8 micro-
satellite of activin type II receptor (ACVR2), although both genes become frameshift-mutated in >80% of
MMR-defective colorectal cancers. To experimentally determine the effect of microsatellite length upon
frameshift mutation in gene-specific sequence contexts, we altered the microsatellite length within
TGFBR2 exon 3 and ACVR2 exon 10, generating A7, A10 and A13 constructs. These constructs were cloned
1 bp out of frame of EGFP, allowing a −1 bp frameshift to drive EGFP expression, and stably transfected
into MMR-deficient cells. Subsequent non-fluorescent cells were sorted, cultured for 7–35 days and har-
vested for EGFP analysis and DNA sequencing. Longer microsatellites within TGFBR2 and ACVR2 showed
significantly higher mutation rates than shorter ones, with TGFBR2 A13, A10 and A7 frameshifts measured
at 22.38 × 10−4, 2.17 × 10−4 and 0.13 × 10−4, respectively. Surprisingly, shorter ACVR2 constructs showed
three times higher mutation rates at A7 and A10 lengths than identical length TGFBR2 constructs but compar-
ably lower at the A13 length, suggesting influences from both microsatellite length as well as the sequence
context. Furthermore, the TGFBR2 A13 construct mutated into 33% A11 sequences (−2 bp) in addition to
expected A12 (−1 bp), indicating that this construct undergoes continual subsequent frameshift mutation.
These data demonstrate experimentally that both the length of a mononucleotide microsatellite and its
sequence context influence mutation rate in defective DNA MMR.

INTRODUCTION

Microsatellites are repetitive DNA sequences consisting of
nucleotide units ranging from 1 to 6 bp and are ubiquitous
throughout the genome (1). The majority of microsatellites
are located in non-coding regions of the genome, but a minor-
ity are present in coding regions (exons) of key growth regu-
ulatory genes such as transforming growth factor beta type II
receptor (TGFBR2), activin type II receptor (ACVR2), phos-
phatase and tensin (PTEN) homolog and Bcl-2-associated X
protein (BAX) (2–6). When DNA mismatch repair (MMR),
the principal enzyme system that directs repair of slippage
mistakes at microsatellite sequences, is defective, frameshift
mutations within alleles of these genes are accumulated and
their protein functions lost (7,8). This mode of gene mutation
is thought to drive the pathogenesis of colorectal cancer and
other tumors with microsatellite instability (MSI) (7–9). For
example, an A10 microsatellite within exon 3 of TGFBR2
and an A8 microsatellite within exon 10 of ACVR2 become
frameshift-mutated (A9 for TGFBR2 and A7 for ACVR2) at
both alleles in 70–90% of colorectal cancers with MSI
(2,4,10). These biallelic frameshift mutations inactivate the
TGFBR2- and ACVR2-encoded protein receptors and allow
tumors to escape the growth-suppressive effects that are

*To whom correspondence should be addressed at: University of Michigan, 3101 Taubman Center, 1500 E. Medical Center Drive, Ann Arbor, MI
48109, USA. Tel: +1 7349364495; Fax: +1 7342323838; Email: jcarethe@umich.edu

© The Author 2010. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oxfordjournals.org
mediated by these two receptors and their intracellular SMAD signaling (4,9,11). We have measured frameshift mutation rates of the TGFBR2 exon 3 A_{10} microsatellite and the ACVR2 exon 10 A_{8} microsatellite and demonstrated that both accumulate −1 bp frameshift mutations in human colorectal cancer cells with DNA MMR defects (hMLH1−/− and hMSH6−/−) after forming heteroduplexes at a constant rate of formation (12). We also observed a higher frameshift mutation rate at the A_{10} microsatellite of TGFBR2 exon 3 compared with the A_{8} microsatellite of ACVR2 exon 10 in MMR-deficient cells (12), which is presumed to be due to the longer microsatellite within TGFBR2. However, ACVR2 has identical length A_{8} polyadenine tracts in exon 3 and exon 10 of its gene, and only its exon 10 microsatellite sequence is mutated in colonic tumors with MSI, indicating exonic selectivity of ACVR2 for frameshift mutation (4). We replicated this observation in a cell model (13) and demonstrated that this exonic selectivity is partially caused by the immediate flanking DNA sequences surrounding both coding microsatellites of ACVR2 (13). Our observations suggest a role of the flanking sequence context for microsatellite mutation with defective DNA MMR in addition to the general observation that longer length microsatellites are more likely to undergo frameshift mutation (8), although this last point has not been experimentally tested for human genes.

Several characteristics of microsatellites have been shown to influence the extent of their instability with defective DNA MMR, using non-coding sequences in mostly yeast or bacterial models, and include: repeat-unit length [e.g. A_{n} versus (GA)_{n} versus (GAC)_{n}], length of the microsatellite (e.g. A_{7} versus A_{13}), base composition (e.g. A_{n} versus G_{n}), sequence context and the degree of ‘perfection’ of the repeated microsatellite; pure repeats are less stable than mixed repeats (e.g. A_{17} is less stable than A_{9}G_{8}) (14–21). However, there are no data regarding frameshift mutation rates on microsatellite length within actual human genes with DNA MMR deficiency, largely because a good model to measure human microsatellites in DNA MMR deficiency, non-fluorescent sequences mutate more frequently in DNA MMR deficiency. To test this, we changed the length of microsatellite sequences within TGFBR2 exon 3 and ACVR2 exon 10 and generated three different constructs (A_{7}, A_{10} and A_{13}). These constructs were cloned into an EGFP plasmid 1 bp out of frame (OF) such that a −1 bp frameshift mutation could be detected by EGFP expression. Observed in real time, we demonstrate that longer microsatellites within TGFBR2 and ACVR2 coding sequences undergo higher frameshift mutation rates in DNA MMR deficiency. We further demonstrate that TGFBR2 containing an A_{13} microsatellite sequence is the most prone to the frameshift mutation when DNA MMR is defective, even over the A_{13} microsatellite constructed for ACVR2. These findings showed experimentally that the length of a coding microsatellite greatly determines its mutation rate with defective DNA MMR. Additionally, by directly comparing identical length ACVR2 and TGFBR2 constructs, we showed that other factors such as the sequence context play a role in determining frameshift mutation rates.

**RESULTS**

**Longer microsatellites within TGFBR2 and ACVR2 coding sequences mutate more frequently in hMLH1−/− cells**

To test whether coding microsatellite length influences frameshift mutation in DNA MMR deficiency, we constructed TGFBR2 and ACVR2 plasmids containing different microsatellite lengths (A_{7}, A_{10} and A_{13}) within exon 3 of human TGFBR2 and exon 10 of human ACVR2 sequences by modifying previously constructed plasmids (12) that were designed to detect a −1 bp frameshift mutation (Table 1) as described in Materials and Methods. In our cell model, a −1 bp frameshift mutation at the microsatellite is detected by EGFP expression because the TGFBR2 and ACVR2 sequences are cloned 1 bp OF of EGFP immediately after the start codon of the EGFP gene in pRESHyg2-EGFP. Mutation resistant (MR) in frame (IF) of the EGFP plasmids and MR OF plasmids were constructed as positive and negative controls, respectively, for each length of microsatellite within each gene (TGFBR2 and ACVR2) by interrupting the microsatellite sequence to prevent frameshift mutation. Positive and negative control plasmids and corresponding experimental plasmids were transfected into hMLH1-deficient (completely DNA MMR-defective) HCT116 cells, and stably transfected cell lines (Supplementary Material, Table S1) were established by hygromycin B selection for mutation analysis. Sequences containing 9, 11 or 12 polyadenines were not used for the establishment of stable cell lines due to the predicted generation of a new in-frame stop codon or a new out-frame start codon which would confound mutation analysis by EGFP expression. At 2 weeks after selection, the proportion of fluorescent cells for each cell line was measured by flow cytometry. All MR IF cell lines containing the various microsatellite lengths within TGFBR2 or ACVR2 showed fluorescence between 93 and 100% (median 99%), demonstrating robust EGFP expression. MR IF cell lines containing TGFBR2 (A_{7}), TGFBR2 (A_{10}) or TGFBR2 (A_{13}) OF sequences or ACVR2 (A_{7}), ACVR2 (A_{10}) or ACVR2 (A_{13}) OF sequences revealed newly fluorescent cells ranging between 0.01 and 4.18% (median 0.31%) net fluorescence over matched counterpart cell lines containing MR OF sequences.

To compare mutation frequencies and rates of cell lines containing TGFBR2 or ACVR2 sequences with different-length microsatellites in DNA MMR deficiency, non-fluorescent cells containing either MR TGFBR2 (A_{7}) OF, TGFBR2 (A_{10}) OF, TGFBR2 (A_{13}) OF, MR TGFBR2 (A_{10}) OF, MR TGFBR2 (A_{13}) OF, MR ACVR2 (A_{7}) OF, ACVR2 (A_{7}) OF, MR ACVR2 (A_{10}) OF, ACVR2 (A_{10}) OF, MR ACVR2 (A_{13}) OF or ACVR2 (A_{13}) OF sequences were sorted and exponentially grown for 7–35 days. At specific time points (days 7, 14, 21, 28 and 35), two cultures of each cell line were analyzed in parallel for EGFP expression by using flow cytometry to detect a −1 bp frameshift mutation. Details of flow cytometry analysis are described previously (12). Representative EGFP histograms at day 21 are shown in Figure 1 for mutation analysis that compares EGFP expression in hMLH1−/− TGFBR2 and hMLH1−/−ACVR2 cells containing different-length microsatellites (A_{7}, A_{10} and A_{13}). hMLH1−/− MR TGFBR2 OF cell lines demonstrated approximately three
to four times less EGFP expression (0.02–0.08%) compared with hMLH1⁻/⁻ MR ACVR2 OF cell lines, which indicates a higher background of EGFP expression of hMLH1⁻/⁻ ACVR2 cell lines. Except for hMLH1⁻/⁻ TGFBR2 OF and ACVR2 OF cell lines containing the A7 microsatellite, all other OF cell lines revealed significantly higher EGFP expression (0.5–6.6%) compared with each MR counterpart cell line (P < 0.05). The longer the microsatellite length was within both TGFBR2 and ACVR2, the higher the proportion of EGFP expression was observed (A13 > A10 > A7) with 46-, 24- and 1.4-fold higher EGFP expression in TGFBR2 OF cell lines and 18-, 15- and 1.2-fold higher EGFP expression in ACVR2 OF cell lines compared with each MR counterpart cell line (Fig. 1). As we observed previously in hMLH1⁻/⁻ TGFBR2 (A10) OF and ACVR2 (A8) OF cells (12), two distinct EGFP-expressing cell populations, M1 and M2, were observed from hMLH1⁻/⁻ TGFBR2 OF and ACVR2 OF cells containing A10 or A13 microsatellite sequences, in which the M2 cells showed brighter EGFP expression compared with the M1 cells (data not shown). However, there was no statistical increase in the proportion of the M2 population in hMLH1⁻/⁻ TGFBR2 (A2) OF and ACVR2 (A2) OF cells when compared with each MR counterpart.

The –1 bp frameshift mutation frequency at each time point was expressed as a fold change using the following formula: (EGFP-positive cells/total live cells from TGFBR2 OF or ACVR2 OF cells)/(EGFP-positive cells/total live cells from MR TGFBR2 OF or MR ACVR2 OF cells). As we observed previously (12), the M2 population from the TGFBR2 (A10), TGFBR2 (A13), ACVR2 (A10) and ACVR2 (A13) sequences in the hMLH1⁻/⁻ cells dramatically accumulated over time, whereas the M1 population showed relatively little change over time (Fig. 2), indicating that the M1 and M2 are distinct populations. The M1 and M2 populations were plotted separately for analysis of mutation frequency. Regarding the M1 population (Fig. 2A), there was no consistent overall increase in the mutation frequency for any construct, but the cells containing longer microsatellites showed a relatively higher mutation frequency in both TGFBR2 and ACVR2 sequences compared with MR constructs (Fig. 2A, insert). Comparing TGFBR2 and ACVR2, the TGFBR2 (A10) sequence showed significantly higher mutation frequency than ACVR2 (A10) sequence after day 7 (P < 0.05) in the M1 population, but there was no significant difference in mutation frequency of the M1 population between TGFBR2 and ACVR2 sequences containing A7 and A13 microsatellites. For the M2 population, the hMLH1⁻/⁻ cells containing longer microsatellites within both TGFBR2 and ACVR2 sequences showed a significantly higher mutation frequency (P < 0.05), at day 35, we observed a 501-, 90- and 3-fold change in TGFBR2 cell lines and a 429-, 129- and 14-fold change in ACVR2 cell lines at A13, A10 and A2 microsatellites, respectively (Fig. 2B). However, overall, there was no significant difference in mutation frequency between TGFBR2 and ACVR2 sequences with A10 and A13 microsatellites for the M2 population. Although the total percentages of EGFP-positive cells are low (~0.03%) in the M2 population, hMLH1⁻/⁻ ACVR2 (A2) OF cells showed significantly higher mutation frequency compared with hMLH1⁻/⁻ TGFBR2 (A2) OF cells (P < 0.05).

TGFBR2 is more susceptible to frameshift mutation at an A13 microsatellite sequence compared with ACVR2

We performed DNA sequencing analysis to confirm that EGFP expression from the M1 and M2 populations was caused by a –1 bp frameshift mutation at the microsatellites of TGFBR2 OF and ACVR2 OF. At day 21 after being plated as non-fluorescent cells, the M1 and/or M2 cell populations of hMLH1⁻/⁻ MR TGFBR2 (A10) OF, TGFBR2 (A7) OF, MR TGFBR2 (A10) OF, TGFBR2 (A10) OF, MR TGFBR2 (A13) OF, TGFBR2 (A13) OF, MR ACVR2 (A7) OF, ACVR2 (A7) OF, MR ACVR2 (A10) OF, ACVR2 (A10) OF, MR ACVR2 (A13) OF and ACVR2 (A13) OF cells were sorted and cultured for ~2 weeks for DNA sequencing analysis. DNA from each cell line was amplified by PCR and then sequenced to assess for frameshift mutation at the microsatellites (Fig. 3). No frameshift mutation was observed in any MR TGFBR2 OF or MR ACVR2 OF cell lines as we expected (data not shown), indicating that EGFP fluorescence in these cells is the background fluorescence. As we observed in hMLH1⁻/⁻ TGFBR2 (A10) OF and ACVR2 (A8) OF cells previously (12), the M2 population of hMLH1⁻/⁻ TGFBR2 and hMLH1⁻/⁻ ACVR2 OF cell lines revealed –1 bp frameshifted

<table>
<thead>
<tr>
<th>EGFP plasmids</th>
<th>Microsatellite sequences A7</th>
<th>A10</th>
<th>A13</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR TGFBR2 IF</td>
<td>A7 GA2C A⇒</td>
<td>A7 CA2G A⇒</td>
<td>A7 GA2C GA2C A⇒</td>
</tr>
<tr>
<td>MR TGFBR2 OF</td>
<td>A7 GA2CA⇒</td>
<td>A7 CA2GA3⇒</td>
<td>A7 GA2CA GA2C A⇒</td>
</tr>
<tr>
<td>TGFBR2 OF</td>
<td>A⇒</td>
<td>A⇒</td>
<td>A⇒</td>
</tr>
<tr>
<td>MR ACVR2 IF</td>
<td>A7 GA2C A⇒</td>
<td>A7 CA2G A⇒</td>
<td>A7 GA2C GA2C A⇒</td>
</tr>
<tr>
<td>MR ACVR2 OF</td>
<td>A7 GA2C A⇒</td>
<td>A7 CA2G A⇒</td>
<td>A7 GA2C GA2C A⇒</td>
</tr>
<tr>
<td>ACVR2 OF</td>
<td>A⇒</td>
<td>A⇒</td>
<td>A⇒</td>
</tr>
</tbody>
</table>

The various TGFBR2 exon 3 and ACVR2 exon 10 sequences were inserted immediately after the translation initiation codon of EGFP gene, being IF of the EGFP gene or 1 bp OF of the EGFP gene in the pRES hyg2-EGFP plasmid. MR plasmids were constructed by interrupting microsatellite sequences to prevent any frameshift mutation. MR IF and MR OF plasmids were used as positive and negative controls for EGFP expression, respectively. OF plasmids are the experimental plasmids to detect a ~1 bp frameshift mutation by EGFP expression. Open reading frames (ORFs), the corresponding positive controls for EGFP expression (MR plasmids), and the frameshift mutation frequencies at each time point are shown in Table 1.
microsatellite sequences (A7 to A6, A10 to A9 and A13 to A12) except hMLH1\(^{--}\) cells that lacked enough M2 population cells to be sorted for DNA sequencing (Fig. 3).

The M2 populations of hMLH1\(^{--}\) TGFBR2 (A10) OF, ACVR2 (A7) OF, ACVR2 (A10) OF and ACVR2 (A13) OF cells revealed \(-1\) bp frameshifted microsatellite without a significant overlap with each corresponding non-frameshifted microsatellite sequence, indicating full mutant (A6, A9 and A12) paired with T6, T9 and T12, respectively) formation that induces bright EGFP expression as observed previously (12). Furthermore, the M2 population of hMLH1\(^{--}\) TGFBR2 (A13) OF cells showed a more extensive frameshift mutation pattern. It showed a higher proportion of \(-1\) bp frameshifted microsatellite sequence (A12) overlapping with other shortened microsatellite sequences (indicated by arrows in Fig. 3), suggesting multiple heteroduplex formation with further frameshifts in the M2 population. The TGFBR2 (A13) OF sequence was the only sequence to show evidence of continuing mutation beyond a \(-1\) bp frameshift full mutation.

On the other hand, no frameshift mutation was observed in the M1 population of either hMLH1\(^{--}\) TGFBR2 (A7) OF or hMLH1\(^{--}\) ACVR2 (A7) OF cells, which was anticipated from their lower mutation frequencies compared with the other OF cells containing longer microsatellites (Figs 1 and 2). As observed previously (12), the M1 population of hMLH1\(^{--}\) TGFBR2 (A10) OF cells revealed a \(-1\) bp
frameshifted microsatellite sequence (A_9) without significant overlap with the WT microsatellite sequence, but the M1 population of hMLH1^{+/−} ACVR2 (A_{10}) OF cells revealed a higher proportion of the WT microsatellite sequence overlapping with an A_9 microsatellite sequence (Fig. 3). Additionally, the M1 population of hMLH1^{+/−} ACVR2 (A_{13}) OF cells revealed a higher proportion of A_{13} microsatellite sequence overlapping with A_9 microsatellite sequence. Different from hMLH1^{+/−} ACVR2 (A_{13}) OF cells, the M1 population of hMLH1^{+/−} TGFBR2 (A_{13}) OF cells revealed a higher proportion of −1 bp frameshifted microsatellite sequence (A_{12}) overlapped with multiple shortened microsatellite sequences (arrows in Fig. 3). This type of sequence data indicate heteroduplex formation in the M1 population of hMLH1^{+/−} TGFBR2 (A_{13}) OF cell line as observed in its M2 population, although the proportion of its overlapped sequences seems to be less compared with the M2 population.

To dissect the overlapping sequences in the M1 and M2 populations of hMLH1^{+/−} TGFBR2 (A_{13}) OF and hMLH1^{+/−} ACVR2 (A_{13}) OF cells, DNA from each cell line was amplified by PCR, subcloned into a TA cloning vector, and single-cell clones were individually sequenced. As suggested by the data in Figure 3, single-DNA clones from the M1 and M2 populations of TGFBR2 (A_{13}) OF cells revealed different microsatellite frameshifts (Table 2). In the M1 population of TGFBR2 (A_{13}) OF cells, the majority of DNA clones (52%) showed a −1 bp frameshift sequence (A_{12}); 22% (5/23) of clones showed an A_{13} sequence. Interestingly, A_{10} (−3 bp frameshifted, 13%) and A_{11} (−2 bp frameshifted, 13%) sequences were also observed in the M1 population of TGFBR2 (A_{13}) OF cells, as indicated in the pooled sequences (arrows in Fig. 3). The M2 population of TGFBR2 (A_{13}) OF cells also revealed a −1 bp frameshift (A_{12}) as the dominant microsatellite (50%); the M2 population showed much less non-frameshifted microsatellite (A_{13}) (4.2%) and more −2 bp frameshifted microsatellite (A_{11}) sequences (33.3%) compared with the M1 population, indicating continuing further deletion frameshift mutation on top of the −1 bp frameshift mutation in the M2 population of TGFBR2 (A_{13}) OF cells (these cells were sorted with the −1 bp frameshifts expressing fluorescence). The TGFBR2 (A_{13}) OF also showed an A_{10} microsatellite sequence (4.2%), and it revealed A_{14} and A_{15} microsatellite sequences.

Figure 2. Mutation frequencies of TGFBR2 exon 3 and ACVR2 exon 10 containing different-length microsatellites in the hMLH1^{+/−} background. Non-fluorescent cells were analyzed for EGFP expression by flow cytometry at 7, 14, 21, 28 and 35 days after being sorted and cultured, and EGFP analysis was performed at each day as shown in Figure 1. Mutation frequency at each time point was expressed as a fold change using the following formula: (EGFP-positive cells/total live cells from TGFBR2 OF or ACVR2 OF cells)/(EGFP-positive cells/total live cells from MR TGFBR2 OF or MR ACVR2 OF cells). The M2 population from the TGFBR2 (A_{10}), TGFBR2 (A_{13}), ACVR2 (A_{10}) and ACVR2 (A_{13}) sequences in the hMLH1^{+/−} background greatly accumulated over time, whereas the M1 population showed relatively little change over time, indicating that the M1 and M2 are distinct populations. Overall, longer microsatellites within both TGFBR2 and ACVR2 sequences showed a higher mutation frequency at the microsatellite in both M1 and M2 populations (A_{13} > A_{10} > A_{7}). Cell lines showing lower mutation frequencies (<30-fold change) were separately plotted in the right panel using a smaller y-axis scale. Data are means from two independent experiments at each time point.
(4.2% each) that were not observed in the M1 population. This indicates frameshift expansion in the TGFBR2 (A13) OF cells, a finding not observed in the A7 and A10 sequences.

Conversely, the majority of DNA clones in the M1 population of hMLH1−/− TGFBR2 (A13) OF cells revealed the A13 microsatellite sequence (68.5%) with small proportions of A10, A12 and A14 sequences (10.5%), whereas the majority of DNA clones in the M2 population of ACVR2 (A13) OF cells revealed A12, A13 sequences as dominant. No frameshift mutation was observed in the M1 populations of hMLH1−/− TGFBR2 (A13) OF and hMLH1−/− ACVR2 (A13) OF cells. The M2 population containing TGFBR2 (A10) OF, ACVR2 (A10) OF and ACVR2 (A13) OF sequences showed −1 bp frameshifted sequence with minimal overlap with the WT sequence, suggesting stable full mutant formation. However, the M2 population of TGFBR2 (A13) OF cells revealed an A12 microsatellite sequence as dominant but overlapped with further frameshifted microsatellite sequences (arrows), suggesting additional heteroduplex formation with continuing frameshifts.

The sequencing analysis indicates that TGFBR2 and ACVR2 sequences mutate differently at longer microsatellites (A13). The M1 populations of both TGFBR2 (A13) OF and ACVR2 (A13) OF sequences indicate heteroduplex formation, but TGFBR2 (A13) OF cells revealed more −1 bp frameshifts. Furthermore, the M2 population of TGFBR2 (A13) OF cells suggests a continuing transition from a full-mutant population to further heteroduplex populations, generating additional shortened frameshifted sequences. These observations suggest that hMLH1−/− TGFBR2 (A13) OF cells are the most prone to frameshift mutation at the microsatellite compared with other OF cell lines in this study.

**The −1 bp frameshift mutation rates at the microsatellite within TGFBR2 and ACVR2 coding sequences are dependent on the length of the microsatellite**

The M2 population was used to calculate the mutation rates at microsatellites of TGFBR2 and ACVR2 sequences by the ‘method of the mean’ as described previously (12). The mutation rate at the microsatellite of TGFBR2 (A13) sequence in hMLH1−/− deficiency was highest: 22.38 × 10⁻² ± 2.50 × 10⁻⁴, but it was not significantly different from that of the ACVR2 (A13) sequence (18.32 × 10⁻⁴ ± 0.52 × 10⁻⁴).
Cells from the M1 and M2 populations of hMLH1−/− TGFBR2 (A13) OF and hMLH1−/− ACVR2 (A13) OF cells were sorted and cultured. DNA from each cell line was amplified by PCR, subcloned and all single-cell clones were individually sequenced to assess for frameshift mutation. The M1 population of both cell lines indicates heteroduplex formation (a mixture of mutant and WT), but TGFBR2 (A13) OF cells revealed more −1 bp frameshifts when compared with ACVR2 (A13) OF cells. In the M2 population, ACVR2 (A13) OF cells revealed −1 bp frameshifts (74%) dominantly, suggesting full mutant formation (A12T12), whereas TGFBR2 (A13) OF cells indicated continuous heteroduplex formation (dominantly A13T13) passing through full mutant formation (A12T12). This suggests that TGFBR2 is more susceptible to frameshift mutation at the A13 microsatellite in the hMLH1−/− background compared with ACVR2.

The lack of significant difference may be caused by a smaller proportion of −1 bp frameshifts in the M2 population of TGFBR2 (A13) OF cells compared with ACVR2 (A13) OF cells (Table 2; 50 versus 79%). As predicted in Figures 1 and 2, significantly higher mutation rates were observed in the hMLH1−/− cells containing longer microsatellites within both TGFBR2 and ACVR2 compared with shorter microsatellites (P < 0.05) (Table 3). The mutation rate at the microsatellite of TGFBR2 (A13) sequence is ~10 times lower than that of TGFBR2 (A13) sequence, whereas it is ~16 times higher than that of TGFBR2 (A7) sequence (Table 3). Mutation at the microsatellite of ACVR2 (A13) sequence is ~2.6 times more frequent than mutation at the microsatellite of ACVR2 (A10) sequence, which is ~16 times higher compared with that at the microsatellite of ACVR2 (A7) sequence. At the A7 microsatellite sequence, ACVR2 showed approximately three times higher −1 bp frameshift mutation rate compared with TGFBR2 (A7) sequence (P < 0.05). Different from the mutation frequency observed in Figure 2B, ACVR2 showed approximately three times higher −1 bp frameshift mutation rate compared with TGFBR2 at A10 microsatellite sequence (P < 0.05). Overall, the longer the length of the coding microsatellite, the higher its mutation rate.

**DISCUSSION**

In this study, we measured and compared mutation rates at varying lengths of microsatellites within actual human coding sequences: TGFBR2 exon 3 and ACVR2 exon 10 in hMLH1−/− colorectal cancer cells. Other studies have shown that the propensity for DNA slippage increased with the length of microsatellite by studying non-coding microsatellite sequences (14,17,20). In addition, by utilizing coding microsatellites within human genes, we demonstrated that the mutation rate at the A10 microsatellite of TGFBR2 exon 3 is significantly higher than the A8 microsatellite of ACVR2 exon 10 in DNA MMR-defective cells, which was hypothesized to be caused by the longer microsatellite length within TGFBR2 (12). Besides the length of a microsatellite, other factors may influence mutation rates at microsatellite sequences such as repeat-unit length, base composition of the microsatellite and sequence context (14–20). Exonic selectivity of ACVR2 for frameshift mutation has been observed in colorectal cancers with MSI (4,10) as well as in DNA MMR-deficient human colorectal cells (13) even though both microsatellites within ACVR2 are identical. Furthermore, we demonstrated that the ACVR2 exonic selectivity is partially caused by the flanking DNA sequences surrounding each microsatellite within ACVR2 (13). This suggests that the sequence context may be a key factor that allows frameshift mutation independent of microsatellite length. The effect of microsatellite length upon frameshift mutation rates has not previously been tested for coding microsatellites. We had hypothesized that longer microsatellites within TGFBR2 and ACVR2 would undergo more frameshift mutation in DNA MMR-deficient cells. To examine this hypothesis, we utilized the DNA MMR-deficient hMLH1−/− cell line in which we transfected TGFBR2 exon 3 and ACVR2 exon 10 sequences containing different-length microsatellites (A7, A10 and A13) with the sequences 1 bp OF to EGFP, which allowed us to observe −1 bp frameshift mutations occurring in real time with the onset of EGFP fluorescence. This experimental design allowed us to compare mutation rates of different-length microsatellites within the same gene, as well as the same microsatellite length across two different genes (i.e. two different sequence contexts).

As we observed in TGFBR2 exon 3 and ACVR2 exon 10 OF cells previously (12), OF cells containing A10 and A13 microsatellites revealed two distinct EGFP populations of mutant cells, an M1 population expressing dim fluorescence and an M2 population expressing bright fluorescence (Fig. 1). As we noted previously, the M2 population accumulated over time, but the M1 population showed little change (Fig. 2).
We observed significantly higher mutation frequencies in \( hMLH1^{-/-} \) cells containing longer microsatellites within \( TGFBR2 \) and \( ACVR2 \) in both M1 and M2 populations (Fig. 2). For the M1 population, which represents heteroduplex intermediates prior to full mutation, we observed a significantly higher mutation frequency in \( hMLH1^{-/-} \) cells containing \( TGFBR2 \) (A10) sequence compared with \( ACVR2 \) (A10) sequence after day 7 (\( P < 0.05 \)) (Fig. 2A), which was confirmed by DNA sequencing (Fig. 3). The M1 population of \( hMLH1^{-/-} \) \( TGFBR2 \) (A10) OF cells revealed a \(-1\) bp frameshift \( (A_9) \) with a small proportion of WT microsatellite sequence \( (A_{10}) \), whereas \( hMLH1^{-/-} \) \( ACVR2 \) (A10) OF cells revealed a higher proportion of A10 microsatellite sequence overlapping with the A9 microsatellite sequence. For the M2 population, \( hMLH1^{-/-} \) \( TGFBR2 \) (A7) OF cells showed significantly higher mutation frequency than \( hMLH1^{-/-} \) \( TGFBR2 \) (A7) OF cells (\( P < 0.05 \)) (Fig. 2B). Except for these two observations, we did not find any significant difference in mutation frequency for the same length of microsatellite between \( TGFBR2 \) exon 3 and \( ACVR2 \) exon 10 sequences in either M1 or M2 populations.

DNA sequencing analysis revealed different frameshift mutation patterns in the cells containing A10 and A13 microsatellites (Fig. 3). As observed previously (12), the M2 populations of \( TGFBR2 \) and \( ACVR2 \) containing the A10 microsatellite demonstrated that they are full mutants, revealing a \(-1\) bp frameshifted sequence \( (A_8 \text{ paired with } T_7) \). The M2 populations of \( TGFBR2 \) and \( ACVR2 \) containing the A13 microsatellite revealed a mixture of frameshift mutations and indicated that the A13 microsatellite sequence is subjected to a continuous and accelerated pattern of frameshift mutation in DNA MMR deficiency. DNA sequencing analysis supports that \( TGFBR2 \) (A13) OF cells are most susceptible to frameshift mutation at this length of microsatellite in the \( hMLH1^{-/-} \) background compared with other OF cell lines.

Mutation rates calculated for \(-1\) bp frameshift mutation within the M2 population confirmed our hypothesis that longer microsatellites within \( TGFBR2 \) and \( ACVR2 \) coding sequences have significantly higher mutation rates in DNA MMR deficiency (Table 3). Compared with the native \( TGFBR2 \) (A10) coding sequence, the modified \( TGFBR2 \) (A13) and \( TGFBR2 \) (A7) sequences showed \( \sim10 \) times higher and \( \sim16 \) times lower mutation rates, respectively. Modified \( ACVR2 \) (A13) and \( ACVR2 \) (A7) sequences showed \( \sim3 \) times higher and \( \sim16 \) times lower mutation rates compared with the modified \( ACVR2 \) (A10) sequence, respectively. The \( TGFBR2 \) (A13) mutation rate is not significantly different from the \( ACVR2 \) (A13) sequence, and the failure to detect a difference is likely caused by characteristics of our cell-model system. Our system does not allow detection of \(-2\) bp frameshifts \( [33\% \text{ of the M2 population from the } TGFBR2 \ (A_{13}) \text{ sequence}, and thus we are unable to calculate such mutation rates. Because of this, we believe that the mutation rate for the \( TGFBR2 \) (A13) sequence is underestimated. In contrast, shorter microsatellite sequences \( (A_{10} \text{ and } A_7) \) allowed us to detect significant differences in mutation rate for identical length microsatellites between \( TGFBR2 \) and \( ACVR2 \) sequences, likely because there was no ongoing frameshift mutation beyond the \(-1\) bp full mutants. In any event, these differences detected with identical length microsatellites suggest other factors that influence frameshift mutation, the most likely being sequence context, which was shown to be a partial cause of exonic selectivity of \( ACVR2 \) for frameshift mutation in defective DNA MMR (13). In \( hMLH1 \) deficiency, substitution of flanking nucleotides surrounding exon 10 microsatellite of \( ACVR2 \) with those surrounding the exon 3 microsatellite of \( ACVR2 \) greatly diminished heteroduplex \( (A_7 \text{ paired with } T_5) \) and full \( (A_7 \text{ paired with } T_7) \) frameshift mutation, whereas substitution of flanking nucleotides from exon 3 with those from the exon 10 enhanced frameshift mutation (13). Additionally, when the immediate surrounding DNA sequences of the A10 microsatellite of \( TGFBR2 \) exon 3 were swapped with DNA sequences from \( ACVR2 \) exon 3, we observed a significant decrease in frameshift mutation at the A10 microsatellite of native \( TGFBR2 \) in the \( hMLH1^{-/-} \) background (Supplementary Material, Fig. S1). All these observations suggest that in addition to microsatellite length, the surrounding sequence context of microsatellite likely regulates frameshift mutation rate in DNA MMR deficiency. As suggested in Figure 4, the sequence context may have more influence on frameshift mutation at shorter microsatellite lengths. At longer microsatellite lengths, the microsatellite length may be the dominant factor influencing frameshift mutation.

In summary, using our cell model in which we measure frameshift mutation in real time, we demonstrate that longer microsatellites within \( TGFBR2 \) exon 3 and \( ACVR2 \) exon 10 coding sequences undergo higher frameshift mutation rates in DNA MMR deficiency than shorter microsatellites. Our observations further suggest that frameshift mutation is influenced by both microsatellite length and its sequence context. Our study has implications for the mutation rates at other
coding microsatellite sequences and for how DNA MMR deficiency drives frameshift mutation.

MATERIALS AND METHODS

Cloning of pIREShyg2-TGFBR2-EGFP and pIREShyg2-ACVR2-EGFP plasmids containing different-length microsatellites

Plasmids pIREShyg2-TGFBR2-EGFP containing portions of exon 3 of TGFBR2 and pIREShyg2-ACVR2-EGFP containing portions of exon 10 of ACVR2 were constructed previously (12). Details of plasmid cloning were previously described (12). With these plasmids containing native coding microsatellite sequences (A10 for TGFBR2 and A8 for ACVR2), pIREShyg2-TGFBR2-EGFP and pIREShyg2-ACVR2-EGFP plasmids containing varying lengths of microsatellites were constructed by changing only microsatellite sequences (A10 to A7 and A13 for TGFBR2 and A8 to A7, A10 and A13 for ACVR2) by using a QuikChange II site-directed mutagenesis kit (Stratagene, La Jolla, CA, USA) (Table 1). For experimental plasmids, TGFBR2 or ACVR2 sequences were cloned 1 bp OF in pIREShyg2-EGFP and thus a −1 bp frameshift mutation at the microsatellite would shift the EGFP gene into the proper reading frame to allow EGFP expression. MR counterpart plasmids were constructed by interrupting microsatellites (A7 to A2GA2CA, A10 to A2CA2GA2CA and A13 to A2GA2CA2GA2CA) using a QuikChange II site-directed mutagenesis kit (Stratagene) to prevent any frameshift mutation. The MR plasmids were placed OF (1 bp) and IF to be used as negative and positive controls for EGFP expression, respectively. Positive colonies were screened, and the correct sequences of TGFBR2 and ACVR2 with different microsatellite lengths were confirmed by DNA sequencing in an ABI 3700 analyzer.

Cell lines, transfection and selection

The human colon cancer cell line, HCT116 (hMLH1−/−), was obtained from American Type Culture Collection (Rockville, MD, USA) and maintained in Iscove’s modified Dulbecco’s medium (Invitrogen Corp., Carlsbad, CA, USA) with 10% fetal bovine serum and penicillin (100 U/ml)/streptomycin (100 µg/ml) (Invitrogen Corp.) as supplements. Cells were transfected with the different pIREShyg2-TGFBR2-EGFP and pIREShyg2-ACVR2-EGFP plasmids (Table 1) by using Nucleofector Kit V (Amaxa, Cologne, Germany) following the manufacturer’s instructions. Selection with hygromycin B (Invitrogen Corp.) commenced at 24 h after nucleasefication to generate stable cell lines. After selection, colonies from each cell line were initially pooled and cultured for mutation analysis. All stable cell lines were confirmed by DNA sequencing.

Mutation analysis by flow cytometry

Details of analysis of mutant cells by flow cytometry were previously described (12). Briefly, 5000 non-fluorescent cells expressing MR TGFBR2 (A7) OF, TGFBR2 (A7) OF, MR TGFBR2 (A10) OF, TGFBR2 (A10) OF, MR TGFBR2 (A13) OF, TGFBR2 (A13) OF, MR ACVR2 (A7) OF, ACVR2 (A7) OF, MR ACVR2 (A10) OF, ACVR2 (A10) OF, MR ACVR2 (A13) OF or ACVR2 (A13) OF were sorted into 24-well plates on a FACS ARIA by using Diva software [Becton Dickinson Immunocytometry Systems (BDIS), San Jose, CA, USA]. During a 7–35 day analysis period, cultures were expanded as required to keep cells in exponential growth. Cells were trypsinized, washed in PBS and resuspended in a total volume of 200 µl of PBS/0.5 µg/ml of propidium iodide (PI) and 3% BSA. Cell suspensions were analyzed on a FACS Calibur with CELL QUEST acquisition and analysis software (BDIS). At specified time points, two cultures were analyzed in parallel. To identify EGFP-positive cells, region (R)1, R2 and R3 were set according to cell size, fluorescence and live cells, respectively. Cells from the gated R1 and R3 and R2 were further plotted on a fluorescence intensity histogram. The population displaying no fluorescence and two distinct EGFP-positive populations were separated. The population with dim EGFP fluorescence was designated M1, and the population with bright EGFP fluorescence was designated M2. The counts of M1 and M2 cells were expressed as percentages of R3 (total live cell number). Different EGFP gates were set to properly distinguish the M1 and M2 populations of hMLH1−/− TGFBR2 and hMLH1−/− ACVR2 cell lines. A lower EGFP gate was set for hMLH1−/− TGFBR2 cell lines compared with hMLH1−/− ACVR2 cell lines.

PCR and DNA sequencing

Total cellular DNA from stable cell lines and M1 and M2 cell populations were PCR-amplified by specific primers 5′-GGTCGTGTTAAAACCTGCTTTCTCAAGTGATTATG-3′ and 5′-TGCCGTCGTTTAAAGGAAGAAGA-3′ for TGFBR2 and 5′-GATCGGCCACCACTGTTAAGAACGC-3′ and 5′-GCTGTGTTAGTTGTACTCCAGTTG-3′ for ACVR2 in a reaction containing the primers, buffer, DNA template, deoxynucleotides and Pfu ultra high-fidelity DNA polymerase (Stratagene). The PCR products were used for DNA sequencing to identify stable cell lines and frameshift mutations at microsatellites of TGFBR2 and ACVR2. In addition, we subcloned PCR-amplified TGFBR2 OF (A13) and ACVR2 OF (A13) DNA fragments from the M1 and M2 cell populations by utilizing a TA cloning vector (Invitrogen Corp.) as per the manufacturer’s protocol. DNA clones were then individually sequenced to determine the prevalence of mutatred and A13 microsatellite sequences of TGFBR2 and ACVR2.

Determination of −1 bp frameshift mutation rates at the microsatellite of TGFBR2 and ACVR2 in MMR-deficient cells

Mutation rates were calculated by the method of the mean developed by Luria and Delbrück (22) as described previously (12). The formula used in the computation is \( \hat{r} = \mu N / (\mu NC) \), where \( \hat{r} \) is the mean number of mutants in a culture, \( C \) the number of parallel cultures, \( \mu \) the mutation rate and \( N \) the number of cells at risk of undergoing a mutation, which Luria–Delbrück assumed to be equal to the final number of cells in a culture. Two parallel cultures were used, and \( \hat{r} \) was estimated as the mean of the number.
of mutants across the two cultures. The total number of cells, \( N \), was based on averaging two cultures. The mutation rates of the M2 cell population (full mutants) were calculated using data from flow cytometry analysis at each time point between day 21 and day 35. Single-mutation rates were then calculated by combining and averaging time-specific mutation rates to minimize the variance of the estimate. Data were expressed as mean ± the standard errors of mean (SEM).

**Statistical analysis**

Mutation frequencies and rates of cell lines were compared by \( t \)-test or one-way ANOVA.

**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at HMG online.

**ACKNOWLEDGEMENTS**

We thank the support of the following shared resources: Moores UCSD Cancer Center Flow Cytometry Shared Resource and Moores UCSD Cancer Center DNA Sequencing Shared Resource.

**Conflict of Interest statement.** None declared.

**FUNDING**

This work was supported by the United States Public Health Service (DK067287 to J.M.C.) and the UCSD Digestive Diseases Research Development Center (DK080506).

**REFERENCES**