

# The nucleotide composition of microsatellites impacts both replication fidelity and mismatch repair in human colorectal cells

Christoph Campregher<sup>1,2</sup>, Theresa Scharl<sup>3,4</sup>, Manuela Nemeth<sup>1,2</sup>, Clemens Honeder<sup>1,2</sup>, Thomas Jascur<sup>5</sup>, C. Richard Boland<sup>5</sup> and Christoph Gasche<sup>1,2,\*</sup>

<sup>1</sup>Christian Doppler Laboratory for Molecular Cancer Chemoprevention and <sup>2</sup>Division of Gastroenterology and Hepatology, Department of Medicine 3, Medical University of Vienna, Vienna, Austria, <sup>3</sup>Department of Statistics and Probability Theory, University of Technology, Vienna, Austria, <sup>4</sup>Department of Biotechnology, University of Natural Resources and Applied Life Sciences, Vienna, Austria and <sup>5</sup>Department of Internal Medicine, Baylor University Medical Center, Dallas, TX, USA

Received October 19, 2009; Revised April 12, 2010; Accepted April 22, 2010

**Microsatellite instability is a key mechanism of colon carcinogenesis. We have previously studied mutations within a (CA)13 microsatellite using an enhanced green fluorescent protein (EGFP)-based reporter assay that allows the distinction of replication errors and mismatch repair (MMR) activity. Here we utilize this assay to compare mutations of mono- and dinucleotide repeats in human colorectal cells. HCT116 and HCT116+chr3 cells were stably transfected with EGFP-based plasmids harboring A10, G10, G16, (CA)13 and (CA)26 repeats. EGFP-positive mutant fractions were quantitated by flow cytometry, mutation rates were calculated and the mutant spectrum was analyzed by cycle sequencing. EGFP fluorescence pattern changed with the microsatellite's nucleotide sequence and cell type and clonal variations were observed in mononucleotide repeats. Replication errors (as calculated in HCT116) at A10 repeats were 5–10-fold higher than in G10, G16 were 30-fold higher than G10 and (CA)26 were 10-fold higher than (CA)13. The mutation rates in hMLH1-proficient HCT116+chr3 were 30–230-fold lower than in HCT116. MMR was more efficient in G16 than in A10 clones leading to a higher stability of poly-G tracts. Mutation spectra revealed predominantly 1-unit deletions in A10, (CA)13 and G10 and 2-unit deletions or 1-unit insertion in (CA)26. These findings indicate that both replication fidelity and MMR are affected by the microsatellite's nucleotide composition.**

## INTRODUCTION

Microsatellite instability (MSI) is detected in approximately 10–15% of colorectal, endometrial and gastric cancers (1,2), a proportion of which are caused by the Lynch syndrome (2,3). In Lynch syndrome, MSI is caused by a defect in the DNA mismatch repair (MMR) system that results from germline mutations in certain MMR genes including hMLH1, hPMS2, hMSH2 or hMSH6 (4). A permanent frameshift mutation in microsatellites is caused by slippage of DNA polymerase followed by a lack of post-replication MMR (5,6). Polymerase slippage happens predominantly in areas

of microsatellites in the eukaryotic genome (7). Most microsatellites are located in non-coding regions, but some of these repeats (typically mononucleotide repeats) are found in coding regions of genes which may be inactivated through frameshift mutations in MSI-positive tumors (8).

The components of the MMR system play a major role in maintaining genetic stability during cell division by correcting replication errors which—if not repaired—would create frameshifts and lead to nonsense mutations. In eukaryotes, homologs of the bacterial MutS- and MutL-MMR proteins form heterodimers with specific roles in the repair of certain types of mismatch errors caused by polymerase slippage (7). MMR is initiated when

\*To whom correspondence should be addressed at: Department of Medicine 3, Division of Gastroenterology and Hepatology, and Christian Doppler Laboratory for Molecular Cancer Chemoprevention, Medical University of Vienna, Waehringer Guertel 18-20, 1090 Vienna, Austria. Tel: +43 1404006154; Fax: +43 1404004724; Email: christoph.gasche@meduniwien.ac.at

complexes of the MutS homologs, either MSH2-MSH6 (MutS $\alpha$ ) or MSH2-MSH3 (MutS $\beta$ ), recognize a mismatch. Eukaryotic DNA polymerases- $\alpha$ , - $\beta$  and - $\delta$  differ in their frequency and specificity of producing frameshift errors, polymerase- $\beta$  being the least accurate enzyme (9). These specificities reflect a major role for eukaryotic polymerases in modulating the integrity of DNA repeats. Polymerase- $\epsilon$  along with polymerase- $\delta$  plays a major role in DNA replication, recombination and repair (10). However, frameshifts and MSI were not significantly increased in exonuclease mutant Pol- $\epsilon^{e/c}$  cells (11).

Microsatellites have a repeat-unit size of 1–6 bp and occupy ~3% of the human genome (12). They consist of 6–30 repeat units and tend to be highly polymorphic (13). A number of characteristics of DNA repeats are known to influence their degree of instability. These include the length of the repeat unit (14), the base composition (15), the number of DNA repeats (16,17), the sequence context (18,19) and the degree of ‘perfection’ of the repeat tract (20). Such studies were based on the use of selective media. When quantitating mutation rates, the presence of preexisting mutations within such assays cannot be precluded.

Previously, we developed a flow cytometry-based assay for the quantitation of frameshift mutations within a (CA)<sub>13</sub> microsatellite (21). This assay could distinguish between actual replication errors and their removal by the DNA MMR system. Here we utilize this assay to compare mutation rates of various mono- and dinucleotides. The aim was to investigate the influence of length, composition and unit type on both the occurrence of replication errors and the removal of such by the MMR.

## RESULTS

### Generation of frameshift-reporter plasmids

The plasmid pIRESHyg2-enhanced green fluorescent protein (EGFP) (21) allows the expression of EGFP under the control of a constitutive cytomegalovirus promoter. Several DNA repeat oligonucleotides were inserted after the translation initiation codon of the EGFP gene shifting it out of its reading frame and resulting in the expression of a truncated peptide without fluorescence. Deletions or insertions within the repetitive sequence may restore the proper reading frame of the EGFP gene and cause expression of EGFP. pIRESHyg2-EGFP was linearized with the restriction enzymes *PmeI* (generating a 3' blunt end) and *AscI* (generating a 5'-GCGC overhang). Compatible DNA repeat oligonucleotides (A10, G10, G16, (CA)<sub>26</sub> and N26) were generated by hybridization of forward and reverse single DNA oligonucleotides with a 5'-GCGC overhang and a 3' blunt end. N26 is a random DNA sequence which served as non-repeat control (5'-GCGCTTAT ACTAAGCGGAAATCGTTA-3'). After ligation, the products were transformed into *StbI2* competent bacteria (Gibco). Amplified plasmids were isolated and sequenced using EGFP-specific primers flanking the DNA repeat sequence.

### Establishment and characterization of frameshift-reporter cell lines

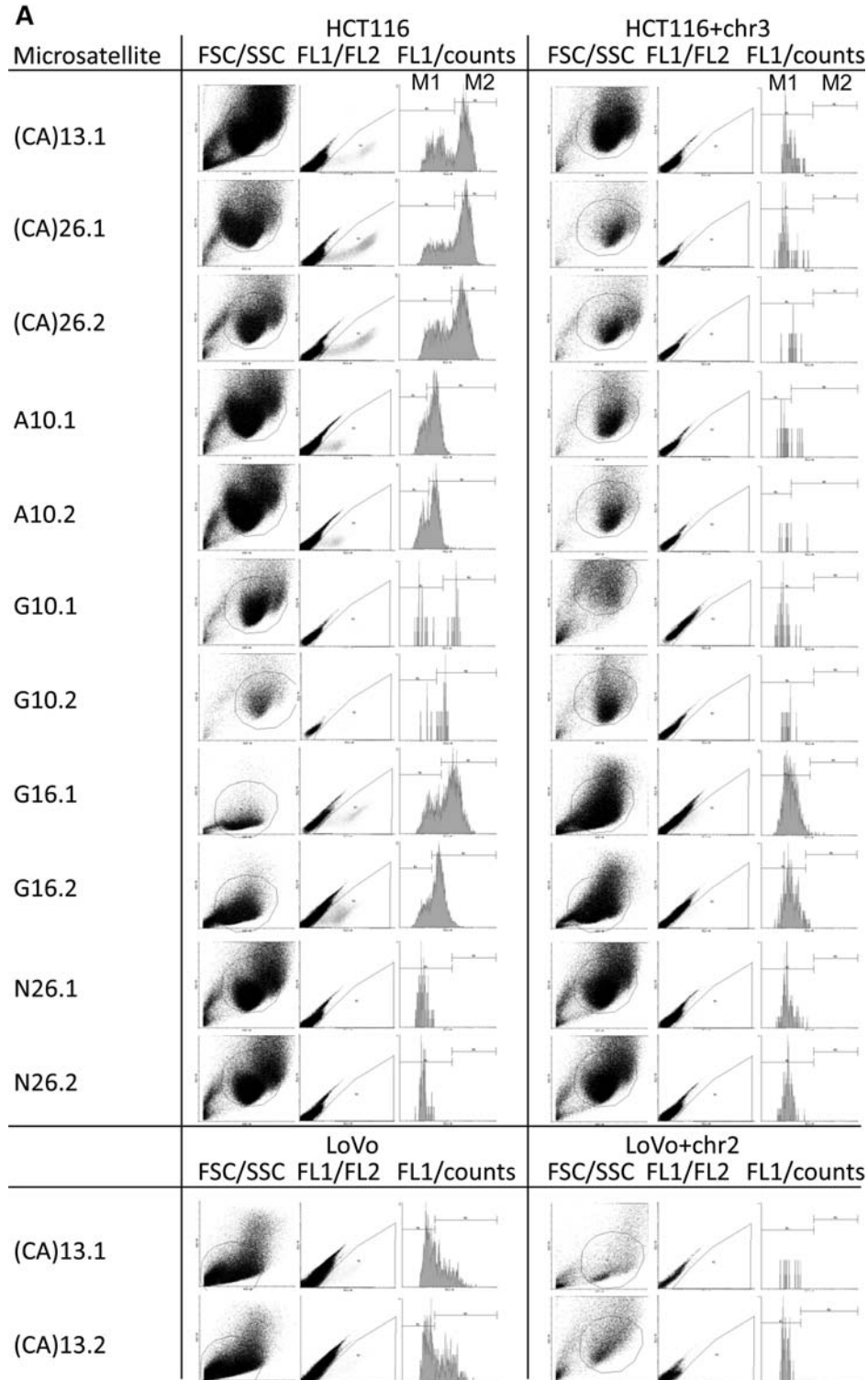
hMLH1-proficient and hMLH1-deficient cells (HCT116+chr3, HCT116) were stably transfected with pIRESHyg2-EGFP-A10,

pIRESHyg2-EGFP-G10, pIRESHyg2-EGFP-G16, pIRESHyg2-EGFP-(CA)<sub>26</sub> DNA repeats and pIRESHyg2-EGFP-N26 similar to our previous experiments using pIRESHyg2-EGFP-(CA)<sub>13</sub> (21). Additionally, hMHS2-proficient and hMSH2-deficient cells (LoVo+chr2, LoVo) were transfected with pIRESHyg2-EGFP-(CA)<sub>13</sub>. Stable clones were characterized by sequencing, Southern blotting and flow cytometry (Supplementary Material, Fig. S1). Out of each transfection, two clones harboring a single insertion of the plasmid were selected and processed further for mutation analyses.

### EGFP fluorescence pattern changes with the nucleotide sequence of the microsatellite insert and the cell type

We utilized a recently developed EGFP-based assay for the detection of frameshift mutations within a (CA)<sub>13</sub> microsatellite (21) and extended this system to various DNA repeat constructs to test whether the composition, length and unit type of DNA repeats are crucial for the frequency of the occurrence of frameshift mutations. Non-fluorescent (M0) cells were sorted into 24-well plates (Supplementary Material, Fig. S2A). EGFP-positive populations were analyzed by flow cytometry after 3–20 days, dependent on the MMR background and proliferation rate. In previous experiments using pIRESHyg2-EGFP-(CA)<sub>13</sub>, the EGFP-positive population consisted of two distinct fractions, M1 and M2, which represent intermediate (correlating with replication errors) and permanent mutant cells (representing the combined effect of replication errors and MMR proficiency) (21). The intensity and distribution of the EGFP-positive populations M1 and M2 were analyzed using Cyflogic flow cytometry data analysis software (Fig. 1). The M1 and M2 fractions displayed diversity in histogram pattern (Fig. 1A; FL1/counts) and fluorescence intensity (Fig. 1B; geometric mean) between different microsatellite sequences but not within clones of the same microsatellite. The mean fluorescence intensity of the M2 population was strongest in HCT116 cells harboring (CA)<sub>26</sub> or (CA)<sub>13</sub> ( $536 \pm 56$  or  $442 \pm 50$ , respectively) followed by G10 ( $151 \pm 88$ ) and G16 ( $117 \pm 78$ ). A10 ( $33 \pm 1$ ) displayed the lowest fluorescence intensity. In some clones (such as HCT116 A10 and HCT116 G16.2), the separation of M1 and M2 fractions was difficult and therefore the results for the M1 fraction have to be considered with caution. LoVo cells harboring a (CA)<sub>13</sub> repeat showed a lower fluorescence intensity than the corresponding HCT116 cells ( $86 \pm 6$  versus  $442 \pm 50$ ). Interestingly, the population pattern (M1 > M2) was also different from the corresponding HCT116 clone (M2 > M1).

In HCT116+chr3 cells, the fluorescence intensity of the M1 populations was comparable and the M2 population was either very small (A10) or not present at all. The non-repetitive control sequence (N26) displayed a very small M1- and no M2 population in both HCT116 and HCT116+chr3 cells. Overall, a good correlation was evident between the fluorescence intensity of the M1 and the according M2 fractions ( $R = 0.932$ ,  $P < 0.0001$ ). These data suggest that the EGFP fluorescence pattern and intensity is altered by the nucleotide composition of the inserted repetitive sequence as well as the type of cell line.



**Figure 1.** EGFP fluorescence pattern and intensity depends on the nucleotide sequence of the microsatellite. Stable transfectants of the various cell clones were sorted for M0 non-fluorescent cells, grown for 10–20 days and were analyzed by flow cytometry on a FACScan (BD Bioscience). (A) The fluorescence pattern of each clone is represented by a histogram (FL1) to quantitate the distinct fluorescence subpopulations M1 and M2. Differences in fluorescence pattern are related to the nucleotide sequence of the microsatellite and the type of cell line. (B) Single cell clones of HCT116 cells were cultured in quadruplicate and the geometric mean of the FL1 intensity was expressed as mean  $\pm$  SD for the M1 and the M2 fractions. A strong variation was observed between different microsatellites but less within single cell clones of the same microsatellite.

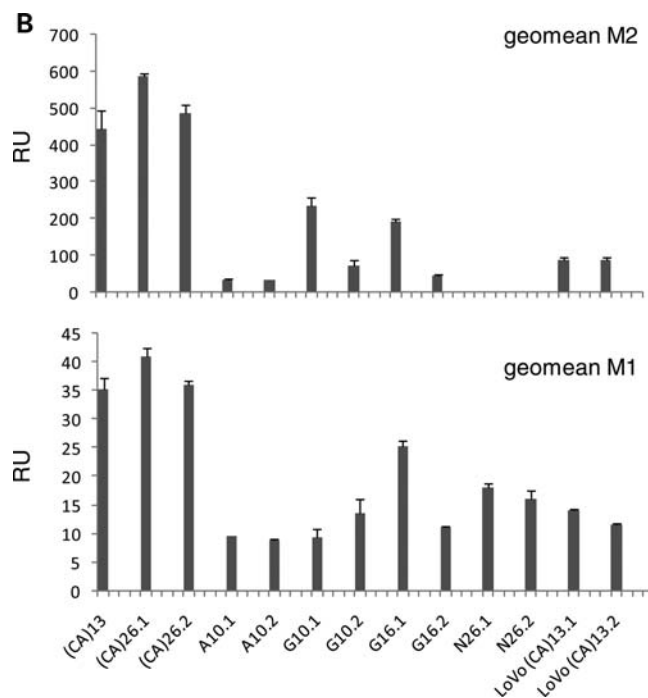


Figure 1. Continued.

### The nucleotide composition and length increase the mutation rate in MMR-deficient cells

Next we compared the change of M1 and M2 mutant fractions over time. M0 cells were sorted, cultured as above and analyzed at early (days 4–8), mid (days 7–12) and late time points (days 10–20) according to the different proliferation rates of clones [LoVo and LoVo+chr2 grow about three times slower than HCT116 (doubling time about 14 h) or HCT16+ hr3 (doubling time about 24 h)]. In most clones, the M1 population showed little change and the M2 population accumulated over time. A comparison between A10 and G10 clones in HCT116 (Fig. 2A) revealed an approximately 4-fold larger M1 fraction in cells harboring the A10 repeat. The corresponding accumulation of M2 cells was also much higher in the A10 clones, while there was little or no change of the M2 population of G10 clones. Comparative mutation rates (Table 1; method of the mean) of A10 were 5–15-fold higher than G10. These data suggest that A10 repeats are more susceptible to replication errors than G10 and, consequently, produce more permanent frameshift mutations in MMR-deficient mammalian cells. The replication fidelity within the G10 tract is, however, significantly compromised when the length of the repeat increases to G16 (Fig. 2B). In fact, the G16 microsatellite is hypermutable with about 3% of M2 cells after 5–6 replication cycles (4 days of culture). Comparative mutation rates (Table 1; method of the mean) of G16 were 20–60-fold higher than G10 (Table 1).

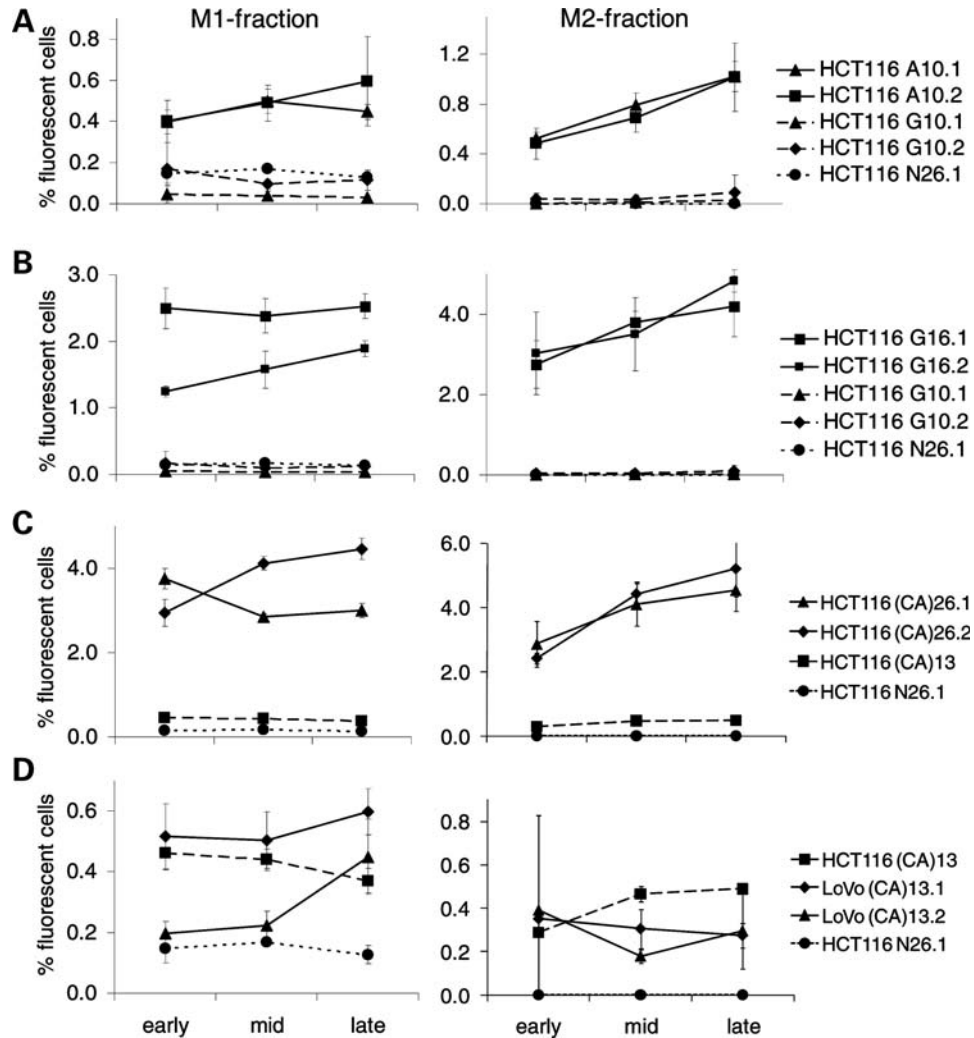
Comparison of (CA)13 and (CA)26 repeats revealed higher mutant fractions for the longer tract both in the M1 and the M2 fraction (Fig. 2C) with an 8-fold difference in mutation rates (Table 1; method of the mean). Changes in mutant fractions in the (CA)13-harboring HCT116 cells were consistent with

our previous data (21). The somewhat lower mutation rate was due to a change in the assumed cloning efficiency (now adjusted to 20%). Despite the obvious differences in EGFP fluorescence pattern of the M1 and M2 population, the size of the M1 and M2 populations was similar in (CA)13-transfected LoVo cells and HCT116 cells (Fig. 2D). Due to the slow growth of LoVo cells, the culture period and conditions were difficult to compare.

Comparative mutation rates of the investigated microsatellites in HCT116 cells were:  $G10 < (CA)13 < A10 < G16 < (CA)26$  ( $P < 0.05$ ). Thus, the stability of microsatellites in MMR-deficient colon cells depends on the nucleotide composition and length of the DNA repeat. Despite the different growth and fluorescence pattern of mutant cells, the mutation rate at the (CA)13 repeat was similar in LoVo and HCT116 cells.

### Mutation spectra of mono- and dinucleotide repeats in MMR-deficient cells

During DNA replication, the formation of insertion/deletion loops within microsatellite sequences leads to a gain or loss of repeat units (22). To analyze the spectrum of the resulting mutations, single cells of the M0, M1 or M2 populations (Supplementary Material, Fig. S2A) were sorted into 96-well plates and grown for 10 days. From each clone, a fluorescent image was taken by inverted fluorescence microscopy (Supplementary Material, Fig. S2B), the DNA was isolated and the microsatellite region was subjected to cycle sequencing (Supplementary Material, Fig. S2C). The cloning efficiency in 96-well plates was 4–25% for M0, 1–21% for M1 and 8–24% for M2 clones. The quality of sequence analysis was good except for G16 clones (which we were unable to



**Figure 2.** Comparison of M1 and M2 populations over time within different microsatellite sequences. Non-fluorescent M0 cells were sorted into 24-well plates and analyzed after a period of 5–20 days (early, mid and late time points) and a change in M1 and M2 fractions was compared between certain clones. (A) Comparison between A10 and G10 clones. (B) Comparison between G10 and G16 clones. (C) Comparison between (CA)13 and (CA)26 clones. (D) Comparison between (CA)13 in HCT116 and LoVo cells. A random non-repeat sequence (N26) served a control.

amplify). Except for (CA)26 repeats, the M0 fraction revealed wild-type sequences only, whereas the intermediate mutant fraction M1 revealed a mixture of wild-type, heteroduplex DNA and definitive mutations (Table 2).

In clones from the M2 fractions, all mutations within the (CA)13, the A10 and G10 microsatellite were 1-unit deletions, whereas both 1-unit insertions as well as 2-unit deletions were seen in (CA)26-harboring HCT116 cells.

Similar to our previous findings, we identified heteroduplex sequences in the M1 fraction specifically in mononucleotide repeats and in the (CA)26 clones (21). In (CA)26 clones, these were 1-unit/2-unit deletion heteroduplexes according to a mixed fluorescence colony phenotype. In sorted M1 cells containing (CA)26 repeats, fluorescence-positive  $-1$  or  $+2$ -unit mutations but no wild-type sequences were observed. This likely reflects the large proportion of (CA)26 mutant cells that are present within the M0 and M1 population. (CA)13 mutations in hMSH2-deficient LoVo cells displayed a similar spectrum as in HCT116. Our data are in line with the expected range of

mutations and demonstrate a good correlation with the fluorescence phenotype of the originating cell (i.e. M0, M1 or M2).

### Repair of insertion/deletion loops by MutL

Similar mutation analyses were performed in clones of HCT116+chr3 that express a functional MutL complex (through transfer of chromosome 3 where hMLH1 is located) but lack hMSH3 to build MutS $\beta$ . In some clones, no M2 cells were detected and, thus, the calculation of the mutation rate was based on the assumption that a single mutant cell was not detected (Table 1). In general, the corresponding clones in HCT116+chr3 displayed significantly lower mutation rates than in HCT116 cells. The highest mutation rate was measured in A10, followed by G16 and (CA)13 (Table 1). The fold reduction in mutation rate (as calculated by the method of the mean) through transfer of hMLH1 (on chromosome 3) was further plotted for clones that displayed definitive mutations in both cell lines (Fig. 3).

**Table 1.** Mutation rates within mono- and dinucleotide repeats in colon epithelial cells

Microsatellite	HCT116		HCT116+chr3		P-value <sup>a</sup>
	ML	MM	ML	MM	
(CA)13	2.0 ± 0.3 (×10 <sup>-4</sup> )	1.9 ± 0.5 (×10 <sup>-4</sup> )	9.0 ± 4.4 (×10 <sup>-6</sup> )	8.9 ± 5.5 (×10 <sup>-6</sup> )	<0.001
(CA)26.1 <sup>b</sup>	1.5 ± 0.2 (×10 <sup>-3</sup> )	1.5 ± 0.4 (×10 <sup>-3</sup> )	<6.4 ± 3.5 (×10 <sup>-6</sup> )	<6.1 ± 5.0 (×10 <sup>-6</sup> )	0.003
(CA)26.2 <sup>b</sup>	1.7 ± 0.2 (×10 <sup>-3</sup> )	1.7 ± 0.4 (×10 <sup>-3</sup> )	<1.1 ± 0.6 (×10 <sup>-6</sup> )	<1.1 ± 0.9 (×10 <sup>-6</sup> )	0.003
A10.1	3.7 ± 0.5 (×10 <sup>-4</sup> )	3.8 ± 1.0 (×10 <sup>-4</sup> )	1.3 ± 0.7 (×10 <sup>-5</sup> )	1.3 ± 1.0 (×10 <sup>-5</sup> )	0.001
A10.2	3.4 ± 0.5 (×10 <sup>-4</sup> )	3.7 ± 1.0 (×10 <sup>-4</sup> )	1.1 ± 0.5 (×10 <sup>-5</sup> )	1.1 ± 0.6 (×10 <sup>-5</sup> )	0.008
G10.1	1.3 ± 0.7 (×10 <sup>-5</sup> )	2.5 ± 1.3 (×10 <sup>-5</sup> )	<7.7 ± 4.2 (×10 <sup>-6</sup> )	<7.3 ± 6.0 (×10 <sup>-6</sup> )	n.s.
G10.2	5.9 ± 3.1 (×10 <sup>-5</sup> )	7.6 ± 4.8 (×10 <sup>-5</sup> )	<1.0 ± 0.5 (×10 <sup>-5</sup> )	<9.6 ± 7.9 (×10 <sup>-6</sup> )	n.s.
G16.1	1.5 ± 0.2 (×10 <sup>-3</sup> )	1.7 ± 0.4 (×10 <sup>-3</sup> )	9.0 ± 3.5 (×10 <sup>-6</sup> )	9.2 ± 4.7 (×10 <sup>-6</sup> )	0.003
G16.2	1.6 ± 0.2 (×10 <sup>-3</sup> )	1.5 ± 0.3 (×10 <sup>-3</sup> )	3.7 ± 2.0 (×10 <sup>-6</sup> )	4.2 ± 3.1 (×10 <sup>-6</sup> )	<0.001
N26.1	<2.6 ± 1.4 (×10 <sup>-6</sup> )	<2.5 ± 2.1 (×10 <sup>-6</sup> )	<3.5 ± 1.9 (×10 <sup>-6</sup> )	<3.3 ± 2.7 (×10 <sup>-6</sup> )	n.s.
N26.2	<2.7 ± 1.5 (×10 <sup>-6</sup> )	<2.6 ± 2.1 (×10 <sup>-6</sup> )	<3.1 ± 1.6 (×10 <sup>-6</sup> )	<2.9 ± 2.4 (×10 <sup>-6</sup> )	n.s.
	LoVo		LoVo + chr2		
(CA)13.1	1.3 ± 0.3 (×10 <sup>-4</sup> )	1.3 ± 0.4 (×10 <sup>-4</sup> )	<4.8 ± 2.6 (×10 <sup>-5</sup> )	<4.5 ± 3.7 (×10 <sup>-5</sup> )	0.009
(CA)13.2	1.0 ± 0.2 (×10 <sup>-4</sup> )	1.3 ± 0.4 (×10 <sup>-4</sup> )	<1.4 ± 0.8 (×10 <sup>-5</sup> )	<1.4 ± 1.1 (×10 <sup>-5</sup> )	0.06

Data are mean ± SEM. Mutation rates are expressed as mutations per microsatellite per generation.

ML, maximum likelihood method; MM, method of the mean.

<sup>a</sup>Between mutation rates (MM) of MMR-deficient (HCT116, LoVo) and MMR-corrected (HCT116+chr3, LoVo+chr2) cells.

<sup>b</sup>Exact mutation rate cannot be established as clones are a mix of wild-type and 1-unit deletions.

n.s. not significant.

As G16 tracts were 5 to 15 times more susceptible to replication errors than A10, the lower mutation rate in G16 indicates that the presence of a functional MutL complex induces a more efficient repair of G16 than A10 insertion/deletion loops.

## DISCUSSION

In the eukaryotic genome, microsatellites are prone to frameshift mutations as a result of polymerase slippage during DNA replication (22). Such events involve insertion or deletion of repeat units. Known factors that influence repeat instability are the length, composition and repeat-unit size of DNA repeats (23). Mononucleotide repeats are the most abundant repeat motifs in humans and are frequently mutated within certain types of cancer such as Lynch syndrome (24). Recently, a frameshift mutation in the A13 repeat within the 3'-untranslated region of epidermal growth factor receptor (EGFR) was shown to be frequently mutated in MSI colon cancers which are associated with increased EGFR expression (25). Poly-(CA) tracts exhibit frequent mutations in non-neoplastic mucosa from patients with ulcerative colitis (26). A proper understanding of such mutational events in humans may help understand the mechanisms that lead to the accumulation of mutations in cancer tissue.

Here, we utilized an EGFP-based frameshift-reporter system to analyze and compare spontaneous mutation rates within mono- and dinucleotide repeats. Plasmids harboring A10, G10, G16, (CA)13 and (CA)26, as well as a non-repetitive random nucleotide sequence (N26) that put the EGFP gene out of the reading frame, were cloned and transfected into hMLH1-deficient and hMLH1-proficient human colon epithelial cells. In addition, our previously studied (CA)13 reporter plasmid (21) was now tested in hMSH2-deficient cells (LoVo). In MMR-deficient cells, significant differences in the mutation rate were observed among these DNA sequences reflecting the different rate of replication errors. As expected, the most

stable sequence was the non-repetitive N26. Despite some clonal variation, the most stable microsatellite was G10 ( $\sim 5.1 \pm 3.1 \times 10^{-5}$ ) with a 5–15-fold lower mutation rate than the A10 tract ( $\sim 3.8 \pm 1.0 \times 10^{-4}$ ). This is somewhat in contrast to previous studies in *S. cerevisiae* and *E. coli* (19,27) but not to mammalian cells (28). When testing mutation rates in our longitudinal studies (days 4–20), no significant change within single clones was observed (data not shown) indicating an equal occurrence of mutations per cell during the culture period.

When increasing the size of the poly-G repeat from a G10 to G16, the mutation rate increased 20–60-fold ( $\sim 1.6 \pm 0.4 \times 10^{-3}$ ), indicating that the length of G-repeats plays a major role on replication fidelity in human cells. We found that poly-(G)10 repeats are more stable than poly-(A)10. In contrast, poly-(G)17 repeats are less stable than poly-(A)17 repeats (29). When analyzing the corresponding mutation rates in hMLH1-proficient HCT116+chr3 cells, the corresponding mutation rates are 20–220-fold lower than in HCT116 which is a result of MMR due to the presence of MutL. This repair efficiency was far better in G16 (Fig. 3) tracts than in A10 or (CA)13 repeats, indicating that the MMR efficiency is affected by the type of the nucleotide repeat. In *S. cerevisiae*, it was also shown that the mismatch correction depends on the type of microsatellite. However, these studies were done with mutants of the MutS $\beta$  components Msh2, Msh3 and Msh6 (30). As HCT116+chr3 are known to lack hMSH3, it will be interesting to test the differential mutation rate in MMR-proficient normal colorectal cells.

Based on physical properties, CA repeats might be considered similar to random sequences in DNA (31), and should in theory exhibit base stacking and base-pair opening kinetics intermediate between the extremes of mononucleotide G or A runs. In fact, the median mutation rate of (CA)17 was lower than that of G17 and higher than A17 (29). Interestingly, CA tracts also form a poly-pyrimidine/poly-purine motif,

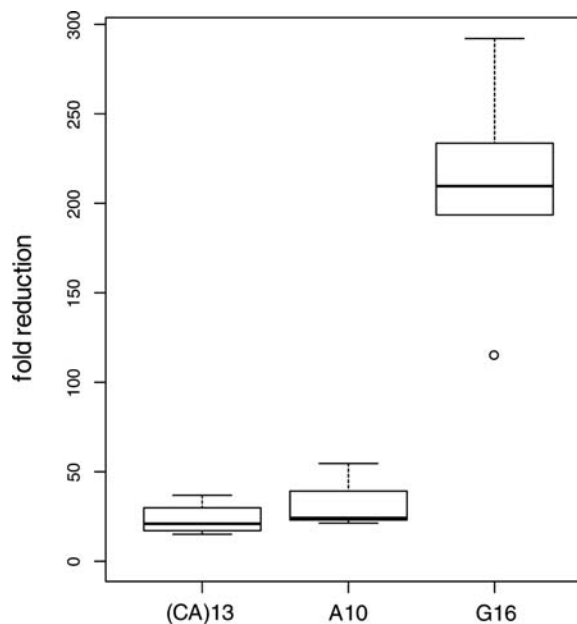
Table 2. Mutation spectrum within DNA microsatellites in colon epithelial cells

	Microsatellite	M0		M1								M2					
		-1	0	-2	-1/-2	-1	0/-1	0	0/+1	+1	+1/+2	+2	-2	-1	0	+1	+2
HCT116	(CA)13		14					8						16			
	(CA)26.1	7 <sup>a</sup>		1	1	8				1			3		1	9	
	(CA)26.2	11 <sup>a</sup>			1	3					2		8			1	
	A10.1		10			3	3	3	2					14			
	A10.2		9	1			1	3						10			
	G10.1		3			3								5			
	G10.2		17			1	3	1						16			
	N26.1		11					5									
	N26.2		8					7									
HCT116+chr3	(CA)13		7					5									
	(CA)26.1		14	3		1	1		5		1	1+4 <sup>b</sup>					
	(CA)26.2		9	3					1			1+1 <sup>b</sup>					
	A10.1		8					5									
	A10.2		9				6	3									
	G10.1		19				1	3									
	G10.2		12					3									
	N26.1		15					13									
N26.2		23					8										
LoVo	(CA)13.1		10				3						8	1			
	(CA)13.2		5					1					7				
LoVochr2	(CA)13.1		8				4	3									
	(CA)13.2		12					6									

No fill color: wild-type or frameshift mutant cells without expected EGFP fluorescence.  
 Green: in-frame mutant cells with expected EGFP fluorescence.  
 M1 light green: Heteroduplex mutant cells with expected partial (dim) EGFP fluorescence.  
<sup>a</sup>Clones turned out to be (CA)25.  
<sup>b</sup>+2/+3 heteroduplexes.

which tends to exhibit Z-DNA characteristics that become greater with length (32,33). The potential for Z-DNA formation of alternating pyrimidine-purine nucleotides [(CG)*n*>(CA)*n*>(TA)*n*] correlates with frameshift potential (33). In our study, mutations at the (CA)26 were 8-fold higher than in (CA)13 and somewhat similar to G16. In fact, cycle sequencing revealed that a large fraction (if not all) of the (CA)26 (non-fluorescent) HCT116 cells harbored actual

(CA)25 microsatellites. Therefore, the exact mutation rate for this clone cannot be calculated. The mutation spectra of (CA)26 (Table 2) showed a large number of definitive mutant cells in the M1 fraction without detecting any cells in the M2 fraction. This indicates that the gate between M1 and M2 was set too stringent, thereby obstructing a proper determination of the mutation rate. It is interesting that the M1 fraction does not exhibit a linear fluorescence pattern



**Figure 3.** Efficiency of MutL in the repair of insertion/deletion loops at different microsatellites. Mutation rates (method of the mean) were compared between hMLH1-deficient HCT116 and hMLH1-proficient HCT116+chr3 and a fold reduction ( $y$ -axis) was calculated. The efficiency to repair G16 mismatches was significantly better than for A10 or (CA)13 ( $P < 0.001$ ).

but rather a clear defined population. A possible explanation could be that cells in the M1 fraction are EGFP positive not because of a failure during replication but rather a transcriptional error caused by RNA polymerase. It is quite interesting that in (CA)26 stretches +2 insertions were as common as -1 deletions, a finding that was not observed with (CA)13, where only -1 deletions were detected (21). Thus, replication errors in longer (CA) stretches produce both insertions and deletions, whereas shorter stretches cause deletions only (16,34). Additionally HCT116 and HCT116+chr3 cells are lacking hMSH3 which recognizes larger loop outs.

By testing the various microsatellites, we identified that different microsatellites change the fluorescence intensity of EGFP. We think this is a methodological issue, specifically with the separation of M1 and M2 fractions or the analysis of rare events (such as the missing M2 fractions in (CA)26 HCT116+chr3 clones). It is likely that this change of fluorescence intensity is due to a peptide-dependent effect of the translated microsatellite sequence at the N-terminus of EGFP. Poly-A and poly-G tracts are translated into poly-Lys or poly-Gly sequences and (CA) repeats into a poly-(His-Thr) sequence. Such differences may limit the use of this mutation reporter assay specifically when it comes to the appropriate setting of the M0/M1 and M1/M2 threshold, which was notoriously more difficult in clones with low EGFP fluorescence. This has to be taken into account when interpreting the data. Additionally, it is difficult to compare fluorescence intensities between LoVo and HCT116 because of a slower proliferation rate of LoVo cells, the culture period and conditions. Interestingly, the clonal variation of the mean fluorescence levels between mononucleotide repeats was greater than between dinucleotide repeats. Therefore, the M1/M2 distinction is

only reliable in poly-CA tracts and the method described previously (21) is not as robust as initially assumed. The control plasmid harboring a random non-repetitive sequence (N26) displayed also a small M1 fraction, but none of these cells were found to be mutant. This most likely represents background noise in the measurement of the M1 population.

The mutation rate analysis with this *in vitro* system has several advantages over the bacterial antibiotic-resistance gene method (34). It can be utilized for a semi-high throughput screening of mutagenic chemicals or of novel compounds when seeking an improvement of replication fidelity (35). It excludes the growth of preexisting mutants by the sorting of fluorescence-negative cells and has an unbiased reporter measurement, whereas antibiotic resistance varies with the antibiotic concentration used. The major value of our model lies in its ability to estimate the relative influence of polymerase errors and MMR proficiency on the rate of frameshift mutations as demonstrated by our data which provide evidence that the nucleotide composition of a microsatellite alters both replication fidelity and MMR efficiency.

## MATERIALS AND METHODS

### Cell culture

HCT116 (hMLH1 mutant) and LoVo (hMSH2 mutant) colorectal cancer cells were obtained from ATCC. HCT116+chr3 (36) (hMLH1 wild-type) and LoVo+chr2 (hMSH2 wild-type) were kind gifts from Drs Boland and Koi. Cells were grown in IMDM (Iscove's modified Dulbecco's medium, GIBCO-Invitrogen, Vienna, Austria) containing 10% fetal bovine serum (Biocrom, Berlin, Germany). The medium for HCT116+chr3 contained additionally 400  $\mu$ g/ml and the medium for LoVo+chr2 contained 700  $\mu$ g/ml Geneticin (G418, GIBCO-Invitrogen). All cells were maintained (so what about G418?) at 37°C, full humidity and 5% CO<sub>2</sub>.

### Plasmid construction and sequence analysis

Using the restriction enzymes *PmeI* and *AscI* (New England Biolabs, Ipswich, GB, USA), the previously established pIRESHyg2-EGFP vector (21) was linearized. Sense and anti-sense oligonucleotides (A10, (CA)26, G10 and N26 = GCGCTTATACTAAGCGGAAATCGTTA as non-DNA repeat control) with a 5'-CGCG overhang to create a compatible site for the linearized pIRESHyg2-EGFP were annealed and directionally cloned into the *PmeI*-*AscI* site of pIRESHyg2-EGFP to generate the plasmids pIRESHyg2-EGFP-(oligonucleotide). The ligation products (pIRESHyg2-EGFP(A)10, pIRESHyg2-EGFP(CA)26, pIRESHyg2-EGFP(G)10 and pIRESHyg2-EGFP(N)26) were transformed into Stb12 (Life Technologies, Rockville, MD, USA) and the cells were grown at 30°C on selection agar plates. Ampicillin-resistant colonies were selected and the correct sequence of the microsatellite insert was confirmed by PCR amplification with 5'-EGFP (5'-CCCACTGCTTACTGGCTTATC G-3') and 3'-EGFP (3'-CCTGAAGTTCATCTGCACCACC-5') and subsequent sequencing with the same primer set on an ABI Prism 310 (Applied Biosystems Inc, Foster City, CA, USA) using the DyEx 2.0 Spin Kit (Qiagen, Hilden, Germany) according to the manufacturer's manual.

### Transfection and stable clone selection

Cells were transfected with the above-described pIREShyg2-EGFP-(oligonucleotide) plasmids using Effectene (Qiagen) according to the manufacturer's manual. Twenty-four hours after transfection, selection with 200  $\mu\text{g/ml}$  hygromycin B (Life Technologies, Gaithersburg, MD, USA) for HCT116 and LoVo or 150  $\mu\text{g/ml}$  hygromycin B and 400  $\mu\text{g/ml}$  or 700  $\mu\text{g/ml}$  geneticin for HCT116+chr3 and LoVo+chr2 cells, respectively. Stable cell clones were picked and analyzed by PCR for EGFP insertion as described above.

### Southern blot analysis

The number of integrated plasmids in the genome was identified by Southern blotting. Total cellular DNA from stably transfected cells that had been grown from single cell colonies was isolated by phenol/chloroform/isoamylalcohol extraction. Twenty micrograms of total DNA were digested with *Bam*HI, *Eco*RV or both (New England Biolabs), resolved on a 0.8% agarose gel and transferred onto a nylon membrane (Hybond N<sup>TM</sup>, Amersham Pharmacia, Braunschweig, Germany) by vacuum blotting (Pharmacia-LKB Vacugene 2016). EGFP cDNA-labeling, hybridization and washing conditions were done as previously described (37).

### Analysis of mutant cells by flow cytometry

Non-fluorescent HCT116, HCT116+chr3, LoVo and LoVo+chr2 frameshift-reporter cells were sorted into 24-well plates on a FACSAria cell sorter by using CloneCyt Plus<sup>TM</sup> software and sorting technology (Becton Dickinson, San Jose, CA, USA). During a 4–20-day growth period, cultures were expanded as required to keep cells in exponential growth. Cells were rinsed with cold  $\text{Ca}^{2+}/\text{Mg}^{2+}$ -free PBS (GIBCO-Invitrogen) and detached with 200  $\mu\text{l}$  Accutase<sup>TM</sup> (PAA Laboratories, Linz, Austria). Fifty microliters of the cell suspension were analyzed on a FACScan (Becton Dickinson) and data were analyzed using Cyflogic flow cytometry data analysis software (Ver. 1.2.1, CyFlo Ltd., Turku, Finland) and cell counts were multiplied by four to quantify the total number of cells per well. Gate settings and analysis were done as described before (21). Briefly, the population displaying no fluorescence was called M0, populations with low fluorescence intensity M1 and those with high fluorescence intensity M2. The counts of M1 and M2 cells were expressed as percentage of R1 (total cell number).

### Analysis of mutation spectrum by sequencing

Single cells of M0, M1 and M2 populations were sorted into 96-well plates and grown for several days to obtain approximately 200 cells per clone. The medium was removed and cells were immediately lysed with 50 mM NaOH and boiled for 10 min at 99°C. The microsatellite locus of the pIREShyg2-EGFP vector was amplified by PCR and further subjected to cycle sequencing as described above to detect the type of frameshift mutations which occurred in the respective microsatellite.

### Statistical analysis and estimation of mutation rates

The mutation rate is defined as the probability of a cell to undergo a mutation in its lifetime, and is expressed per microsatellite per cell per generation. We used two different approaches for estimating the mutation rate: (i) the Luria–Delbruck method (38) and (ii) a maximum likelihood method (39,40). The Luria–Delbruck method is moment-based, whereby the mutation rate is estimated as a function of the sample mean of the number of mutants. In the maximum likelihood approach, a probability-generating function for the number of mutant cells is used to obtain the distribution of mutants as a function of the mutation rate. This likelihood function for the observed data is calculated and maximized for the mutation rate. In both of these methods, the number of cells that undergo a mutation during their lifetime (we will refer to these as definitively mutated cells) is estimated. Then, the mutation rate was obtained by dividing the number of definitively mutated cells by the total number of cells that could potentially have undergone a mutation. Data from M2 cells from the last day of analysis were used to calculate the mutation rate by the method of the mean and the maximum likelihood approaches (41). A cloning efficiency of 20% was considered for the estimation of the mutation rates. In order to assess the difference between cell lines, the mutation rates of clone replicates were compared by the Welch two-sample *t*-test. The Lea–Coulson method of the median was used to calculate the mutation rates of single clones as it is independent of the number of clone replicates. The resulting *P*-values were adjusted for multiple testing using the Benjamini–Hochberg method to control the false discovery rate.

### AUTHOR'S CONTRIBUTIONS

Christoph Campregher (study design, acquisition of data, analysis and interpretation of data, drafting of the manuscript and statistical analysis), Theresa Scharl (statistical analysis), Manuela Nemeth and Clemens Honeder (acquisition of data), Thomas Jascur and C. Richard Boland (material support and critical revision of the manuscript) and Christoph Gasche (study design, interpretation of data, critical revision of the manuscript for important intellectual content, study supervision and obtained funding).

### SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

### ACKNOWLEDGEMENTS

We thank Drs Rick Boland, Minoru Koi and Ajay Goel (Baylor University Medical Center, Dallas) for HCT116+chr3 and LoVo+chr2 cells, Victoria Heinrich for performing Southern blot analysis, and Drs Martin Willheim and Guenther Hofbauer for cell sorting (Core Unit Zellsortierung, Medical University of Vienna). The financial support by the Federal Ministry of Economy, Family and Youth and the National Foundation for Research, Technology and Development is gratefully acknowledged. The funders had no role in

study design, data collection and analysis, decision to publish or preparation of the manuscript.

*Conflict of Interest statement.* None declared.

## FUNDING

This study was supported by the Austrian Science Fund (FWF grant P18270) and the Christian Doppler Research Association. Funding to pay the Open Access charge was provided by FWF.

## REFERENCES

- Boland, C.R., Thibodeau, S.N., Hamilton, S.R., Sidransky, D., Eshleman, J.R., Burt, R.W., Meltzer, S.J., Rodriguez-Bigas, M.A., Fodde, R., Ranzani, G.N. and Srivastava, S. (1998) A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.*, **58**, 5248–5257.
- Ionov, Y., Peinado, M.A., Malkhosyan, S., Shibata, D. and Perucho, M. (1993) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature*, **363**, 558–561.
- Peinado, M.A., Malkhosyan, S., Velazquez, A. and Perucho, M. (1992) Isolation and characterization of allelic losses and gains in colorectal tumors by arbitrarily primed polymerase chain reaction. *Proc. Natl Acad. Sci. USA*, **89**, 10065–10069.
- Eshleman, J.R. and Markowitz, S.D. (1996) Mismatch repair defects in human carcinogenesis. *Hum. Mol. Genet.*, **5**(Spec No.), 1489–1494.
- Streisinger, G. and Owen, J. (1985) Mechanisms of spontaneous and induced frameshift mutation in bacteriophage T4. *Genetics*, **109**, 633–659.
- Henderson, S.T. and Petes, T.D. (1992) Instability of simple sequence DNA in *Saccharomyces cerevisiae*. *Mol. Cell Biol.*, **12**, 2749–2757.
- Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E. and Inouye, M. (1966) Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday. *Cold Spring Harb. Symp. Quant. Biol.*, **31**, 77–84.
- Sharan, S.K. and Bradley, A. (1997) Identification and characterization of a microsatellite marker within murine Brca2 gene. *Mamm. Genome*, **8**, 79.
- Kunkel, T.A. (1986) Frameshift mutagenesis by eucaryotic DNA polymerases in vitro. *J. Biol. Chem.*, **261**, 13581–13587.
- Pursell, Z.F., Isoz, I., Lundstrom, E.B., Johansson, E. and Kunkel, T.A. (2007) Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science*, **317**, 127–130.
- Albertson, T.M., Ogawa, M., Bugni, J.M., Hays, L.E., Chen, Y., Wang, Y., Treuting, P.M., Hedde, J.A., Goldsby, R.E. and Preston, B.D. (2009) DNA polymerase epsilon and delta proofreading suppress discrete mutator and cancer phenotypes in mice. *Proc. Natl Acad. Sci. USA*, **106**, 17101–17104.
- Subramanian, S., Mishra, R.K. and Singh, L. (2003) Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biol.*, **4**, R13.
- Ellegren, H. (2000) Heterogeneous mutation processes in human microsatellite DNA sequences. *Nat. Genet.*, **24**, 400–402.
- Sia, E.A., Jinks-Robertson, S. and Petes, T.D. (1997) Genetic control of microsatellite stability. *Mutat. Res.*, **383**, 61–70.
- Eckert, K.A. and Yan, G. (2000) Mutational analyses of dinucleotide and tetranucleotide microsatellites in *Escherichia coli*: influence of sequence on expansion mutagenesis. *Nucleic Acids Res.*, **28**, 2831–2838.
- Yamada, N.A., Smith, G.A., Castro, A., Roques, C.N., Boyer, J.C. and Farber, R.A. (2002) Relative rates of insertion and deletion mutations in dinucleotide repeats of various lengths in mismatch repair proficient mouse and mismatch repair deficient human cells. *Mutat. Res.*, **499**, 213–225.
- Wierdl, M., Dominska, M. and Petes, T.D. (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics*, **146**, 769–779.
- Kondrashov, A.S. and Rogozin, I.B. (2004) Context of deletions and insertions in human coding sequences. *Hum. Mutat.*, **23**, 177–185.
- Harfe, B.D. and Jinks-Robertson, S. (2000) Sequence composition and context effects on the generation and repair of frameshift intermediates in mononucleotide runs in *Saccharomyces cerevisiae*. *Genetics*, **156**, 571–578.
- Boyer, J.C., Hawk, J.D., Stefanovic, L. and Farber, R.A. (2008) Sequence-dependent effect of interruptions on microsatellite mutation rate in mismatch repair-deficient human cells. *Mutat. Res.*, **640**, 89–96.
- Gasche, C., Chang, C.L., Natarajan, L., Goel, A., Rhee, J., Young, D.J., Arnold, C.N. and Boland, C.R. (2003) Identification of frame-shift intermediate mutant cells. *Proc. Natl Acad. Sci. USA*, **100**, 1914–1919.
- Kunkel, T.A. and Bebenek, K. (2000) DNA replication fidelity. *Annu. Rev. Biochem.*, **69**, 497–529.
- Sia, E.A., Kokoska, R.J., Dominska, M., Greenwell, P. and Petes, T.D. (1997) Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. *Mol. Cell Biol.*, **17**, 2851–2858.
- Bubb, V.J., Curtis, L.J., Cunningham, C., Dunlop, M.G., Carothers, A.D., Morris, R.G., White, S., Bird, C.C. and Wyllie, A.H. (1996) Microsatellite instability and the role of hMSH2 in sporadic colorectal cancer. *Oncogene*, **12**, 2641–2649.
- Yuan, Z., Shin, J., Wilson, A., Goel, S., Ling, Y.H., Ahmed, N., Dopeso, H., Jhaver, M., Nasser, S., Montagna, C. et al. (2009) An A13 repeat within the 3'-untranslated region of epidermal growth factor receptor (EGFR) is frequently mutated in microsatellite instability colon cancers and is associated with increased EGFR expression. *Cancer Res.*, **69**, 7811–7818.
- Brentnall, T.A., Crispin, D.A., Bronner, M.P., Cherian, S.P., Hueffed, M., Rabinovitch, P.S., Rubin, C.E., Haggitt, R.C. and Boland, C.R. (1996) Microsatellite instability in nonneoplastic mucosa from patients with chronic ulcerative colitis. *Cancer Res.*, **56**, 1237–1240.
- Jackson, A.L., Chen, R. and Loeb, L.A. (1998) Induction of microsatellite instability by oxidative DNA damage. *Proc. Natl Acad. Sci. USA*, **95**, 12468–12473.
- Burns, D.P. and Temin, H.M. (1994) High rates of frameshift mutations within homo-oligomeric runs during a single cycle of retroviral replication. *J. Virol.*, **68**, 4196–4203.
- Boyer, J.C., Yamada, N.A., Roques, C.N., Hatch, S.B., Riess, K. and Farber, R.A. (2002) Sequence dependent instability of mononucleotide microsatellites in cultured mismatch repair proficient and deficient mammalian cells. *Hum. Mol. Genet.*, **11**, 707–713.
- Gragg, H., Harfe, B.D. and Jinks-Robertson, S. (2002) Base composition of mononucleotide runs affects DNA polymerase slippage and removal of frameshift intermediates by mismatch repair in *Saccharomyces cerevisiae*. *Mol. Cell Biol.*, **22**, 8756–8762.
- Hud, N.V. and Plavec, J. (2003) A unified model for the origin of DNA sequence-directed curvature. *Biopolymers*, **69**, 144–158.
- Rich, A., Nordheim, A. and Wang, A.H. (1984) The chemistry and biology of left-handed Z-DNA. *Annu. Rev. Biochem.*, **53**, 791–846.
- Freund, A.M., Bichara, M. and Fuchs, R.P. (1989) Z-DNA-forming sequences are spontaneous deletion hot spots. *Proc. Natl Acad. Sci. USA*, **86**, 7465–7469.
- Twerdi, C.D., Boyer, J.C. and Farber, R.A. (1999) Relative rates of insertion and deletion mutations in a microsatellite sequence in cultured cells. *Proc. Natl Acad. Sci. USA*, **96**, 2875–2879.
- Gasche, C., Goel, A., Natarajan, L. and Boland, C.R. (2005) Mesalazine improves replication fidelity in cultured colorectal cells. *Cancer Res.*, **65**, 3993–3997.
- Koi, M., Umar, A., Chauhan, D.P., Cherian, S.P., Carethers, J.M., Kunkel, T.A. and Boland, C.R. (1994) Human chromosome 3 corrects mismatch repair deficiency and microsatellite instability and reduces N-methyl-N'-nitro-N-nitrosoguanidine tolerance in colon tumor cells with homozygous hMLH1 mutation. *Cancer Res.*, **54**, 4308–4312.
- Gasche, C., Chang, C.L., Rhee, J., Goel, A. and Boland, C.R. (2001) Oxidative stress increases frameshift mutations in human colorectal cancer cells. *Cancer Res.*, **61**, 7444–7448.
- Luria, S.E. and Delbruck, M. (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, **28**, 491–511.
- Sarkar, S., Ma, W.T. and Sandri, G.H. (1992) On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. *Genetica*, **85**, 173–179.
- Lea, D.E. and Coulson, C.A. (1949) The distribution of the numbers of mutants in bacterial populations. *J. Genet.*, **49**, 264–285.
- Natarajan, L., Berry, C.C. and Gasche, C. (2003) Estimation of spontaneous mutation rates. *Biometrics*, **59**, 555–561.