

## Relative rates of insertion and deletion mutations in a microsatellite sequence in cultured cells

CHRISTINE D. TWERDI\*, JAYNE C. BOYER\*, AND ROSANN A. FARBER\*†‡§

\*Department of Pathology and Laboratory Medicine, †Curriculum in Genetics and Molecular Biology, and ‡Lineberger Comprehensive Cancer Center, CB 7525, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

Communicated by Clyde A. Hutchison III, University of North Carolina, Chapel Hill, NC, January 13, 1999 (received for review November 23, 1998)

**ABSTRACT** A cell culture system has been used to determine the relative rates of insertions and deletions of integral numbers of dinucleotide repeats in a microsatellite sequence. A plasmid was constructed that contained 17 repeats of poly(dC-dA)·poly(dG-dT) near the 5' end of a bacterial neomycin-resistance (*neo*) gene, such that the *neo* gene was translated in the (+1) reading frame. The plasmid was introduced into mismatch-repair-proficient and mismatch-repair-deficient mammalian cell lines. Rates of mutation to resistance to the neomycin analogue G418 were measured, and the nature of the mutations was determined. The mutations were all gains or losses of integral numbers of repeats, and mutations involving a single repeat greatly predominated over those involving multiple repeats. The data obtained from these studies were compared with results previously obtained with cells transfected with a similar plasmid in which the sequence of the oligonucleotide insert placed the *neo* gene in the (–1) reading frame. This experimental design made it possible to make direct comparisons between insertions and deletions of a single repeat unit. A significant excess of insertions over deletions was found in both repair-proficient and repair-deficient cell lines, although the few mutations involving more than two repeats were deletions.

Microsatellites are simple repetitive DNA sequences that are widely dispersed throughout the eukaryotic genome and are estimated to exist in copies numbering in the tens of thousands (1). The short tandem nucleotide repeats that constitute a microsatellite vary between one and six base pairs (bp) in length, and the entire sequence of a repeat tract is typically less than 100 bp long (2).

A high fraction of microsatellites are highly polymorphic; i.e., the numbers of tandem repeats in a given microsatellite sequence often vary within populations (2–4). These polymorphisms have served as valuable markers in the construction of human genetic maps (5), and they are useful in identity testing (6) and in population genetics studies (7). The polymorphic nature of microsatellite DNA results from frameshift mutations, usually involving the gain or loss of integral numbers of repeat units. It is generally thought that the mechanism for the occurrence of these mutations is slipped-strand mispairing during the process of DNA replication (7–10). Details of the mechanism are under active study in a number of laboratories.

Microsatellite instability has been found in most tumors of patients with hereditary nonpolyposis colorectal cancer (HNPCC) (11). Most of these patients have germ-line mutations in genes involved in mismatch repair (reviewed in ref. 12). This type of instability is also found in a significant fraction of sporadic tumors of certain types (12).

It has been suggested that microsatellite sequences have increased in length during primate evolution (13). This idea initially met with skepticism (14), but subsequent population studies have supported the hypothesis that microsatellites have tended to undergo a higher frequency of insertions than deletions (15–18).

The study reported here was designed to determine directly the relative rates of insertion and deletion mutations in a typical microsatellite sequence in somatic cells. A cell culture system, previously described in ref. 19, has been used to detect frameshift mutations within a poly(dC-dA)·poly(dG-dT), or CA-repeat, microsatellite sequence. The system is based on a plasmid vector into which various microsatellite sequences can be inserted, resulting in the disruption of the reading frame of an adjacent bacterial neomycin-resistance (*neo*) gene. The plasmid integrates into the genome of the recipient cells, and cells that have frameshift mutations in the repeat sequence that restore the reading frame of the *neo* gene can be selected for their ability to grow in the presence of the neomycin analogue G418. We have carried out fluctuation analysis to determine spontaneous rates of mutation to G418 resistance in an immortalized, nontumorigenic mouse fibroblast line with normal mismatch repair and in a human colorectal cancer line lacking activity of the *hMLH1* mismatch repair gene. The types and rates of mutation for an insert in which the *neo* gene is in the (+1) reading frame have been determined. These data have been compared with those obtained with cells containing the same sequence in the (–1) reading frame (19, 20). The rates of mutation involving single repeat units are significantly higher for insertions than for deletions.

### MATERIALS AND METHODS

**Cell Culture.** CAK-Stu3 cells were obtained from R. M. Liskay (Oregon Health Sciences University). CAK-Stu3 is a thymidine-kinase-deficient derivative of the immortalized CAK (21) mouse cell line. These cells are nontumorigenic (22) and have normal mismatch-repair activity (19). CAK-Stu3 cells were maintained in Eagle's MEM supplemented with 10% defined calf serum (HyClone). These cells were previously referred to as CAK cells (19), and they are abbreviated as such in this report.

H6 is a subclone (23) of the human colorectal cancer cell line HCT116 (24), which lacks mismatch-repair activity as the result of the absence of a normal copy of the *hMLH1* gene (25). These cells were provided by B. Vogelstein (The Johns Hopkins University). H6 cells were cultured in Dulbecco's modified Eagle's medium (DMEM).

Abbreviations: CA-repeat, poly(dC-dA)·poly(dG-dT); hyg<sup>R</sup>, hygromycin-resistant; G418<sup>R</sup>, G418-resistant.

§To whom reprint requests should be addressed at: Department of Pathology and Laboratory Medicine, University of North Carolina, CB 7525 Brinkhous-Bullitt Bldg., Chapel Hill, NC 27599-7525. e-mail: rfarber@med.unc.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

All cells were maintained at 37°C in a 5% CO<sub>2</sub>/95% air atmosphere.

**Plasmid Construction.** The pCT1 plasmid (Fig. 1) contains a bacterial hygromycin-resistance (*hyg*) gene and a fusion gene consisting of most of the herpes simplex virus thymidine kinase (*tk*) gene at the 5' end and a bacterial neomycin-resistance (*neo*) gene at the 3' end. A synthetic oligonucleotide containing a (CA)<sub>17</sub> repeat was inserted at an *Aat*III restriction site upstream of the *neo* gene, such that the downstream portion of the fusion gene is read out of frame and no functional protein is produced. The microsatellite insert in pCT1 is 55 bp in length, which puts the *neo* gene into the (+1) reading frame.

**Transfection.** pCT1 plasmid DNA, linearized with *Hind*III, was introduced by electroporation into the CAK cells. Transfected cells were selected on the basis of their ability to grow in medium containing hygromycin B (150 µg/ml).

**Fluctuation Tests.** Subcultures were established from each hygromycin-resistant (*hyg*<sup>R</sup>) clone (100 cells per subculture for CAK cells and 5 cells per subculture for H6 cells) and expanded to >1 × 10<sup>6</sup> cells. Cells with mutations that restored the reading frame of the *neo* gene were selected in medium containing 500–600 µg/ml G418 (Geneticin; GIBCO/BRL). Cells from subcultures of transfected CAK clones were plated at a density of 2–5 × 10<sup>5</sup> cells per 100-mm Petri dish; cells from subcultures of the transfected H6 clones were plated at densities of 2 × 10<sup>3</sup> or 2 × 10<sup>4</sup> cells per dish. G418 was added at the time of plating and with medium exchanges every 4–5 days for approximately 2 weeks. One G418-resistant (G418<sup>R</sup>) colony was isolated from each subculture and grown to confluence in a 25-cm<sup>2</sup> tissue culture flask for subsequent DNA extraction and PCR analysis. Plates were stained with Giemsa stain, and G418<sup>R</sup> colonies were counted. Colony counts were adjusted for the cloning efficiencies of cells from each respective subculture in nonselective medium. Mutation rates were calculated with the Lea and Coulson (26) method of the median. Rates determined for clones containing two copies of the microsatellite were divided by 2 to obtain the mutation rate per microsatellite tract.

Statistical analyses were carried out as described by Wierdl *et al.* (27). Mutation rates for each subculture were calculated by substitution of the numbers of mutant colonies per subculture for the median number of mutant colonies in the Lea and Coulson equation (26). These values were ranked for cells

containing the microsatellite insert in both reading frames.  $\chi^2$  analysis was then carried out on the numbers of subcultures with the microsatellite in the (+1) reading frame that ranked above the median of these values versus the numbers of subcultures of this type that ranked below the median.

**PCRs.** DNA from isolated *hyg*<sup>R</sup> and G418<sup>R</sup> clones was extracted by "salting out" of proteins and precipitation with ethanol (28). PCR amplification was performed with [ $\alpha$ -<sup>32</sup>P]dCTP in the reaction mixture, as previously published (19), to confirm the presence of plasmid DNA in *hyg*<sup>R</sup> clones and to detect changes in length of the microsatellite insert in G418<sup>R</sup> clones. Temperature cycling conditions were as follows: initial 7-min denaturation at 94°C; 27 cycles of 1 min at 94°C, 2 min at 55°C, and 1 min at 72°C; final 6-min extension at 72°C. Amplified products were separated on denaturing 6% polyacrylamide gels (1,500 V, 2 hr). Gels were dried and exposed to Kodak BioMax film overnight at room temperature for visualization of PCR products.

G418<sup>R</sup> clones derived from two CAK *hyg*<sup>R</sup> clones contained multiple copies of the pCT1 plasmid. There were two PCR products from each of the G418<sup>R</sup> clones derived from these clones, one of the same size as the original microsatellite insert and one with an insertion or deletion mutation. (Because G418 resistance is dominant, mutation of only one microsatellite sequence resulted in G418<sup>R</sup> clones.) Phosphorimage analysis was used to estimate the ratio of parental to mutant bands.

**DNA Sequencing.** DNA from G418<sup>R</sup> clones was sequenced by using the Thermo Sequenase radiolabeled-terminator cycle sequencing kit (Amersham Life Science). Reactions were performed according to the manufacturer's protocol for sequencing PCR products.

For sequencing of PCR products from clones containing multiple copies of the microsatellite insert, the mutant PCR product was separated from the parental-type PCR product by polyacrylamide gel electrophoresis. The mutant band of interest was excised from the dried gel and suspended in 100 µl of water to extract the DNA. Sequencing was then performed with the kit described above.

## RESULTS

Because the scheme that we use for determination of microsatellite mutation rates is a frameshift-reversion assay, it is not possible to detect both insertions and deletions of a single repeat unit beginning with the same microsatellite sequence. If the microsatellite places the *neo* gene in the (–1) reading frame, a deletion of a single dinucleotide repeat will correct the reading frame, but an insertion of a single repeat will not. The converse is true for a sequence that places the *neo* gene in the (+1) reading frame. We have, therefore, analyzed the rates and types of mutations obtained for a (CA)<sub>17</sub>-repeat with a *neo* gene in the (+1) frame. These data are compared with results obtained previously from a similar construct in the (–1) frame (CAK-pRTM2) (19).

pCT1 plasmid DNA was linearized and introduced by electroporation into CAK cells; this plasmid is diagrammed in Fig. 1. Transfected cells were selected for resistance to hygromycin. DNA from *hyg*<sup>R</sup> clones was analyzed by PCR to confirm the presence of the CA-repeat. Fluctuation analysis was performed to determine rates of mutation to G418 resistance; these data are shown in Table 1. Results are given for four independent *hyg*<sup>R</sup> clones containing the pCT1 plasmid. Also included in Table 1 are the mutation rates from CAK-pRTM2, as determined by the Lea and Coulson method of the median (26). CAK-pCT1 mutation rates were based on the numbers of G418<sup>R</sup> colonies in 10 or 11 subcultures from each transfected clone; these numbers were each corrected for the cloning efficiency of that subculture. Two of the CAK-pCT1 clones contained two copies each of the microsatellite/*neo* construct. Because reversion to G418 resistance is dominant, a mutation

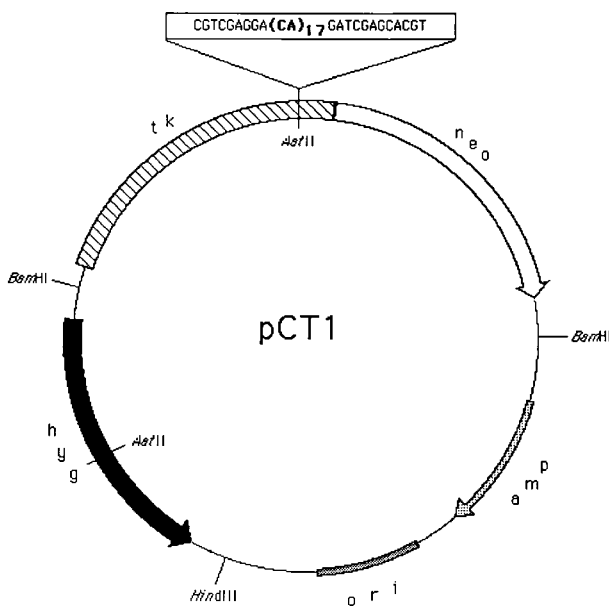


FIG. 1. pCT1 plasmid. Boldface characters within the boxed sequence represent the microsatellite.

Table 1. Rates of microsatellite mutation to G418 resistance in CAK *hyg<sup>R</sup>* clones with the *neo* gene in the (+1) (CAK-pCT1) or (-1) (CAK-pRTM2) reading frame

<i>hyg<sup>R</sup></i> clone no.		<i>hyg<sup>R</sup></i> clone no.	
Mutation rate		Mutation rate*	
CAK-pCT1-6	$1.5 \times 10^{-5}$	CAK-pRTM2-1	$7.5 \times 10^{-6}$
-8	$4.4 \times 10^{-5}$	-2	$1.4 \times 10^{-5}$
-9	$8.1 \times 10^{-5}\dagger$	-6	$1.3 \times 10^{-5}$
-11	$8.0 \times 10^{-6}\dagger$		

Rates are expressed as mutations per cell per generation.

\*Data from ref. 19, recalculated by the Lea and Coulson method of the median (26).

†Clones containing two copies of the microsatellite sequence; mutation rates in these clones have been divided by 2 to determine rates per microsatellite sequence.

in only one copy of the microsatellite is necessary to produce a resistant clone. The mutation rates in these clones were divided by 2, to obtain mutation rates per microsatellite sequence. Mutation rates in clones with the *neo* sequence in the (+1) reading frame were significantly higher than those with the *neo* sequence in the (-1) reading frame, as determined by  $\chi^2$  analysis of mutation rates in every subculture ( $\chi^2 = 6.2$ ,  $P < 0.05$ ).

Results of PCR analysis were used to determine what types and sizes of mutations occurred in individual G418<sup>R</sup> clones. Forty-one clones (one per subculture from the fluctuation tests reported in Table 1) were analyzed. PCR products were subjected to electrophoresis in polyacrylamide gels, adjacent to PCR products from parental *hyg<sup>R</sup>* clones that were not exposed to G418. Fig. 2 shows the results of PCR analysis on eight G418<sup>R</sup> clones isolated from the CAK-pCT1-*hyg<sup>R</sup>*-8 transfectant. The PCR product in the parental clones was 151 bp in length, and the products from all of the G418<sup>R</sup> revertants shown were 153 bp long. As indicated in Table 2, all clones tested had either an insertion or deletion within the microsatellite sequence that was a multiple of 2 bp compared with the control product, and all of the observed shifts would be expected to restore the reading frame of the *neo* gene. (Data on the frequencies of different types of mutations from CAK-pRTM2 cells are also given in Table 2.) Previous results from sequencing of the PCR products in CAK-pRTM2 G418<sup>R</sup> revertants indicated that the length differences detected by PCR were the result of insertions and deletions of integral numbers of dinucleotide-repeat units (19). PCR products from representative CAK-pCT1 G418<sup>R</sup> revertants were also sequenced. These included three +2-bp revertants, two -4-bp revertants, and the single -10-bp and -16-bp revertants. In each instance, the length differences were confirmed to have resulted from the gain or loss of integral numbers of CA-repeat units.

The overall rates of insertion and deletion mutations in all revertants were compared. The frequencies of revertant clones in each subculture were multiplied by the fraction of all clones derived from the same parental cell line (CAK-pCT1 or CAK-pRTM2) that contained insertion mutations, and fluctuation analysis was carried out using these frequencies. Separate calculations were done for the rates of deletion mutations. The median rates of insertion and deletion for G418<sup>R</sup> clones derived from each *hyg<sup>R</sup>* transfectant are given in Table

Table 2. Numbers of base pairs inserted or deleted from microsatellite sequence in CAK cells with the *neo* gene in the (+1) or (-1) reading frame

(+1) Reading frame		(-1) Reading frame	
No. of revertants	Difference in length from parental cells, bp*	No. of revertants	Difference in length from parental cells, bp†
36	+2	14	+4
3	-4	12	-2
1	-10	4	-8
1	-16	1	-16
		1	-20
Total	41	32	

\*Approximately half of these clones contain two copies of the microsatellite sequence, one of which is mutant.

†Data from ref. 19.

3. The rates of insertion mutations in both lines combined were significantly higher than the rates of deletion mutations in both lines ( $\chi^2 = 6.3$ ,  $P < 0.05$ ). As can be seen from the data in Table 3, this difference is almost entirely attributable to 2- to 10-fold higher rates of insertion than deletion in the CAK-pCT1 cells.

The vast majority of the G418<sup>R</sup> revertants from the CAK-pCT1 cells had insertions of a single repeat (36 of 41 clones). The types of mutations identified in revertants from CAK-pRTM2 clones were more heterogeneous; a slight excess of these mutants contained deletions, but nearly half of these clones contained insertions of two repeat units. On the basis of studies of yeast and human cancer cell lines deficient in mismatch repair, the microsatellite mutations that occur most frequently are those involving a single repeat unit (refs. 23, 29, and 30, and data presented below on H6 cells). To determine whether insertions or deletions of single repeats are more common in repair-proficient cells, we compared the rates of insertion of a single repeat in CAK-pCT1 cells with the rates of deletion of a single repeat in CAK-pRTM2. The frequencies of G418<sup>R</sup> clones were each multiplied by the fraction of clones containing +2-bp or -2-bp mutations, respectively. These numbers were then used to calculate mutation rates. The median rates are given in Table 4. The difference between rates of insertion of one CA-repeat in CAK-pCT1 and the rates of deletion of one repeat in CAK-pRTM2 was highly significant ( $\chi^2 = 8.0$ ,  $P < 0.01$ ).

Since the results on CAK cells reflect both occurrence and repair of different types of microsatellite mutations, we carried out a similar analysis on a subclone [H6 (23)] of the human colorectal cancer cell line HCT116, which lacks a functional copy of the *hMLH1* mismatch-repair gene (25). These results should reflect only differences in the rates of occurrence of various types of mutations in the absence of mismatch repair. H6 cells were transfected with pCT1 DNA, and analysis of the rates and types of microsatellite mutations was carried out as described above for CAK cells. All 26 of the G418<sup>R</sup> revertants analyzed by PCR contained insertions of a single CA-repeat. In H6 cells containing the pRTM2 plasmid with the *neo* gene in the (-1) reading frame, 75 of 84 mutation events were deletions of a single repeat; all of the others, except for one anomalous insertion of a single repeat, were insertions of two repeats (20).

The rates of insertion of a single repeat in H6-pCT1 cells were calculated and are given in Table 5, along with the rates of deletion of a single repeat in H6-pRTM2, which were based on the data obtained previously (20). The H6-pCT1 single-repeat insertion rates were significantly higher than the H6-pRTM2 deletion rates ( $\chi^2 = 9.6$ ,  $P < 0.01$ ). The overall microsatellite mutation rates for H6-pCT1 clones were identical to the rates of insertion of a single repeat, since that is the only type of mutation that was observed in these experiments.

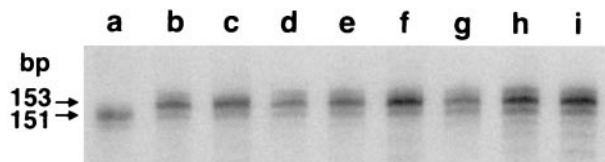


FIG. 2. PCR products from G418<sup>R</sup> mutants of CAK-pCT1-8. Lane a, parental *hyg<sup>R</sup>* cells; lanes b-i, G418<sup>R</sup> clones.



Table 3. Rates of all insertions and deletions in CAK revertant clones

CAK-pCT1 [(+1) reading frame]			CAK-pRTM2 [(-1) reading frame]*		
hyg <sup>R</sup> clone no.	Insertion rate	Deletion rate	hyg <sup>R</sup> clone no.	Insertion rate	Deletion rate
6	$1.3 \times 10^{-5}$	$1.6 \times 10^{-6}$	1	$3.3 \times 10^{-6}$	$7.1 \times 10^{-6}$
8	$3.9 \times 10^{-5}$	$7.1 \times 10^{-6}$	2	$7.9 \times 10^{-6}$	$8.5 \times 10^{-6}$
9	$7.3 \times 10^{-5\dagger}$	$9.8 \times 10^{-6\dagger}$	6	$7.5 \times 10^{-6}$	$7.5 \times 10^{-6}$
11	$1.2 \times 10^{-5\dagger}$	$7.2 \times 10^{-6\dagger}$			

Rates are expressed as mutations per cell per generation.

\*Based on data from ref. 19.

†Clones containing two copies of the microsatellite sequence; mutation rates in these clones have been divided by 2 to determine rates per microsatellite sequence.

The overall mutation rates in the H6-pRTM2 were slightly higher than the rates of deletion of a single repeat (20). These mutation rates were also significantly higher in H6-pCT1 cells than in H6-pRTM2 cells ( $\chi^2 = 9.6$ ,  $P < 0.01$ ).

## DISCUSSION

For a (CA)<sub>17</sub>-repeat in mismatch-repair-proficient cultured mouse CAK cells and in repair-deficient human H6 colorectal cancer cells, we have found that the rates of insertion of integral numbers of repeat units within the microsatellite are significantly higher than the rates of deletion. To make an unbiased comparison using a frameshift reversion assay, we studied cells in which the selectable reporter gene (*neo*) was in either the (+1) or the (-1) reading frame, as the result of the presence of different numbers of bases in the upstream oligonucleotide-insert sequence containing the microsatellite. Since it seemed *a priori* likely that the smallest mutation of a given type (i.e., insertion or deletion) would be the most common within a given line, comparison of +2-bp with -4-bp mutations in cells with the *neo* sequence in the (+1) reading frame might be skewed in favor of the +2-bp mutations, whereas there might be skewing in favor of -2-bp mutations if only cells with the *neo* sequence in the (-1) reading frame were analyzed.

The types of mutations detected in repair-proficient cells with the *neo* sequence in the (-1) reading frame were relatively heterogeneous. On the other hand, there was a striking preponderance of insertions of single repeat units in the cells that carried the *neo* sequence in the (+1) reading frame. Insertions of a single repeat were significantly more common in CAK-pCT1 cells than were deletions of a single repeat in CAK-pRTM2 cells. In general, insertions were significantly more common than deletions among the combined mutants derived from these two lines. Although there were more deletion mutants than insertion mutants among the CAK-pRTM2 cells, the smallest insertion of an integral number of repeats that could restore the reading frame of the *neo* gene (+4 bp) was the single largest class of mutants. The smallest possible insertion (+2 bp) was, by far, the most common type of mutation in CAK-pCT1 cells.

Table 4. Rates of single-repeat insertions in CAK-pCT1 clones and single-repeat deletions in CAK-pRTM2 clones

hyg <sup>R</sup> clone no.	Rate of 2-bp insertions	hyg <sup>R</sup> clone no.	Rate of 2-bp deletions*
CAK-pCT1-6	$1.3 \times 10^{-5}$	CAK-pRTM2-1	$2.5 \times 10^{-6}$
-8	$3.9 \times 10^{-5}$	-2	$5.2 \times 10^{-6}$
-9	$7.0 \times 10^{-5\dagger}$	-6	$6.3 \times 10^{-6}$
-11	$6.5 \times 10^{-6\dagger}$		

Rates are expressed as mutations per cell per generation.

\*Based on data from ref. 19.

†Clones containing two copies of the microsatellite sequence; mutation rates in these clones have been divided by 2 to determine rates per microsatellite sequence.

In the mismatch-repair-deficient H6 cells, nearly all of the mutations involved a single CA-repeat unit. The overall microsatellite mutation rates in the cells with the *neo* gene in the (+1) reading frame (H6-pCT1) were significantly higher than in the cells with the *neo* gene in the (-1) reading frame (H6-pRTM2); therefore, the rates of single-repeat insertions in H6-pCT1 cells were significantly higher than the rate of single-repeat deletions in H6-pRTM2 cells. (We did not make a direct statistical comparison of all insertions vs. all deletions in all of the H6 cells, as we did in CAK cells, since there were no deletion mutants in H6-pCT1 cells.)

On the basis of the observations on repair-deficient cells, it can be concluded that mutations of a single repeat occur most frequently in both reading frames and that the rate of occurrence of insertions of a single repeat is significantly higher than the rate of occurrence of deletions of a single repeat. The findings on the repair-proficient CAK cells reflect a combination of which types of mutations occur and which types are eliminated by the mismatch-repair system. These data suggest that single-repeat insertions may also be repaired less efficiently than single-repeat deletions. In both CAK and H6 cells, all of the mutational events involving more than two repeats were deletions. These larger deletions may balance the more common small insertions, resulting in a tendency for the average length of a given microsatellite to be maintained.

There was considerable variation in microsatellite mutation rates among the CAK-pCT1 hyg<sup>R</sup> clones, although the rates in all of these clones were lower than the overall mutation rates of the CAK-pRTM2 clones. We previously noted variation of this type among H6-pRTM2 clones (20). We have suggested the possibility that these differences are related to the location in the genome into which the plasmid integrates. Another possibility is that there is promoter methylation in some clones; however, preliminary data from experiments using 5-azacytidine in attempts to "reactivate" potentially inactive *tk* promoters in these cells suggest that methylation is not the cause of this variation (unpublished data). We believe that it is more likely that this variation is the result of differences in the sequence contexts into which the plasmids have integrated in independent transfected clones.

Relative frequencies of insertion and deletion mutations have been studied in similar simple-repeat sequences in lower organisms. Levinson and Gutman (31) analyzed CA-repeats in

Table 5. Rates of single-repeat insertions in H6 clones with the *neo* gene in the (+1) (H6-pCT1) reading frame and single-repeat deletions in H6 clones with the *neo* gene in the (-1) (H6-pRTM2) reading frame

hyg <sup>R</sup> clone no.	Rate of 2-bp insertions	hyg <sup>R</sup> clone no.	Rate of 2-bp deletions*
H6-pCT1-1	$6.7 \times 10^{-3}$	H6-pRTM2-A1	$2.5 \times 10^{-3}$
-2	$3.2 \times 10^{-3}$	-A3	$1.4 \times 10^{-4}$
-7	$7.5 \times 10^{-3}$	-B5	$1.1 \times 10^{-4}$

Rates are expressed as mutations per cell per generation.

\*Based on data from ref. 20.

M13 bacteriophage DNA in *Escherichia coli* and found a strong bias in favor of deletions (a ratio of seven deletions to one insertion for 11 repeats and five insertions to one deletion for 20 repeats). The same authors have pointed out, however, that eukaryotic microsatellites might be expected to behave differently, because there is strong selection pressure on bacteria to keep the genome size to a minimum (9).

The pattern of mutations is more complicated in the yeast *Saccharomyces cerevisiae* (32). The rate of insertion of single repeats increases with tract length, but the rate of deletion of single repeats does not. In a wild-type strain, insertions are more common than deletions within tracts 33, 99, and 105 bp in length, whereas deletions outnumber insertions in a 51-bp tract. The rates of occurrence of larger deletions (-4 to -62 bp) are higher in 51-bp and longer tracts than in 15-bp or 33-bp tracts. In an *msh2* strain there is no bias toward insertions or deletions except in a 99-bp tract, in which insertions of a single repeat are predominant.

Previous efforts to determine the relative rates of insertions and deletions in higher eukaryotes have been based primarily on population studies of meiotic events. More gains than losses have been found in an AAAG-repeat that is highly polymorphic in swallows, with allele sizes varying between 36 and 185 repeats (16). Amos *et al.* (15) collected data on germ-line mutations in human CA-repeats and combined them with previously published data on CA- and GATA-repeats and a single polymorphic CTG-repeat (33). They found a significant bias in favor of insertions. A study of allele-frequency distributions for CA-repeats in the Centre d'Étude du Polymorphisme Humain (CEPH) database (5) showed skewing in favor of longer alleles; the results of computer simulation suggested that the underlying model for germ-line mutation in these sequences is likely to be asymmetric, in the direction of increases in allele sizes (18).

The studies cited above on vertebrates have all been of germ-line mutation rates, whereas we have studied microsatellite mutations in cultured somatic cells. There are also other data suggesting a bias in cultured human cells toward insertions. Weber and Wong (33) reported a ratio of 25 increases to 6 decreases in length in lymphoblastoid cells. This finding suggests that cultured cells can be an appropriate model for the study of germ-line mutations, because this bias is similar to what was found for germ-line mutations by the same group, when combined with additional data from the literature (15). The extent to which mouse cells are an appropriate model for microsatellite mutation studies in human cells may not be evident, since simple-repeat sequences are particularly stable in the mouse germ line; however, it has been reported that a trinucleotide repeat is somatically unstable in transgenic mice (7), and we have previously demonstrated that the presence of the (CA)<sub>17</sub>-repeat increases the rate of reversion of the *neo* gene in mouse CAK cells by 100-fold (19).

Most of the studies reporting higher rates of insertion than deletion have involved CA-repeats. The report on mutations in human lymphoblastoid cells included both GATA- and CA-repeats. The distribution was heavily skewed toward insertions (33). It is not clear whether mutations in both types of repeats were equally skewed, but most of these mutations were in the tetranucleotide repeats. There have been reports of bias in favor of deletions in humans in other microsatellites. In a study of eight tetranucleotide-repeats of various sequence composition and one pentanucleotide-repeat in a large number of human meioses, 23 mutational events were identified, among which there was a small net loss of repeats (17). Poly(A) sequences in colorectal cancer cell lines with various degrees of microsatellite instability show a striking imbalance of deletions over insertions (34).

We are aware of no earlier reports of an experimental study on the relative rates of insertions and deletions in cultured mammalian cells. The data reported here argue that there is a

significant bias in favor of insertions in a (CA)<sub>17</sub>-repeat, both in the presence and in the absence of mismatch-repair activity. It will be of considerable interest to determine whether there is a similar bias in other microsatellites, including longer CA-repeat tracts and tracts composed of mononucleotide and tetranucleotide repeats.

We thank Marsha G. Hanford for assistance with calculations and Dr. Thomas D. Petes for helpful comments. This work was supported by National Institutes of Health Grant CA63264.

1. Hamada, H., Petrino, M. G. & Kakunaga, T. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6465-6469.
2. Tautz, D. (1989) *Nucleic Acids Res.* **17**, 6463-6471.
3. Litt, M. & Luty, J. A. (1989) *Am. J. Hum. Genet.* **44**, 397-401.
4. Weber, J. L. & May, P. E. (1989) *Am. J. Hum. Genet.* **44**, 388-396.
5. Dib, C., Fauré, S., Fizames, C., Samsón, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., *et al.* (1996) *Nature (London)* **380**, 152-154.
6. Sprecher, C. J., Puers, C., Lins, A. M. & Schumm, J. W. (1996) *BioTechniques* **20**, 266-276.
7. Djian, P. (1998) *Cell* **94**, 155-160.
8. Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E. & Inouye, M. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31**, 77-84.
9. Levinson, G. & Gutman, G. A. (1987) *Mol. Biol. Evol.* **4**, 203-221.
10. Sia, E. A., Jinks-Robertson, S. & Petes, T. D. (1997) *Mutat. Res.* **383**, 61-70.
11. Aaltonen, L. A., Peltomäki, P., Leach, F. S., Sistonen, P., Pylkkänen, L., Mecklin, J.-P., Järvinen, H., Powell, S. M., Jen, J., Hamilton, S. R., *et al.* (1993) *Science* **260**, 812-816.
12. Eshleman, J. R. & Markowitz, S. D. (1995) *Curr. Opin. Oncol.* **7**, 83-89.
13. Rubinsztein, D. C., Amos, W., Leggo, J., Goodburn, S., Jain, S., Li, S.-H., Margolis, R. L., Ross, C. A. & Ferguson-Smith, M. A. (1995) *Nat. Genet.* **10**, 337-343.
14. Ellegren, H., Primmer, C. R. & Sheldon, B. C. (1995) *Nat. Genet.* **11**, 360-362.
15. Amos, W., Sawcer, S. J., Feakes, R. W. & Rubinsztein, D. C. (1996) *Nat. Genet.* **13**, 390-391.
16. Primmer, C. R., Ellegren, H., Saino, N. & Møller, A. P. (1996) *Nat. Genet.* **13**, 391-393.
17. Brinkmann, B., Klintschar, M., Neuhuber, F., Hühne, J. & Rolf, B. (1998) *Am. J. Hum. Genet.* **62**, 1408-1415.
18. Farrall, M. & Weeks, D. E. (1998) *Am. J. Hum. Genet.* **62**, 1260-1262.
19. Farber, R. A., Petes, T. D., Dominska, M., Hudgens, S. S. & Liskay, R. M. (1994) *Hum. Mol. Genet.* **3**, 253-256.
20. Hanford, M. G., Rushton, B. C., Gowen, L. C. & Farber, R. A. (1998) *Oncogene* **16**, 2389-2393.
21. Farber, R. A. & Liskay, R. M. (1974) *Cytogenet. Cell Genet.* **13**, 384-396.
22. Thomassen, D. G. & DeMars, R. (1982) *Cancer Res.* **42**, 4054-4063.
23. Parsons, R., Li, G.-M., Longley, M. J., Fang, W., Papadopoulos, N., Jen, J., de la Chapelle, A., Kinzler, K. W., Vogelstein, B. & Modrich, P. (1993) *Cell* **75**, 1227-1236.
24. Brattain, M. G., Fine, W. D., Khaled, M., Thompson, J. & Brattain, D. E. (1981) *Cancer Res.* **41**, 1751-1756.
25. Papadopoulos, N., Nicolaidis, N. C., Wei, Y.-F., Ruben, S. M., Carter, K. C., Rosen, C. A., Haseltine, W. A., Fleischmann, R. D., Fraser, C. M., Adams, M. D., *et al.* (1994) *Science* **263**, 1625-1629.
26. Lea, D. E. & Coulson, C. A. (1949) *J. Genet.* **49**, 264-285.
27. Wierdl, M., Greene, C. N., Datta, A., Jinks-Robertson, S. & Petes, T. D. (1996) *Genetics* **143**, 713-721.
28. Miller, S. A., Dykes, D. D. & Polesky, H. F. (1988) *Nucleic Acids Res.* **16**, 1215.
29. Strand, M., Prolla, T. A., Liskay, R. M. & Petes, T. D. (1993) *Nature (London)* **365**, 274-276.
30. Kahn, S. M., Klein, M. G., Jiang, W., Xing, W. Q., Xu, D. B., Perucho, M. & Weinstein, I. B. (1995) *Carcinogenesis* **16**, 1223-1228.
31. Levinson, G. & Gutman, G. A. (1987) *Nucleic Acids Res.* **15**, 5323-5338.
32. Wierdl, M., Dominska, M. & Petes, T. D. (1997) *Genetics* **146**, 769-779.
33. Weber, J. L. & Wong, C. (1993) *Hum. Mol. Genet.* **2**, 1123-1128.
34. Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D. & Perucho, M. (1993) *Nature (London)* **363**, 558-561.