Somatic Mutation Rates and Specificities at TC/AG and GT/CA Microsatellite Sequences in Nontumorigenic Human Lymphoblastoid Cells

Suzanne E. Hile, Guang Yan, and Kristin A. Eckert

The Jake Gittlen Cancer Research Institute, The Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033

ABSTRACT

We have examined mutational events at TC/AG microsatellites, the second most abundant dinucleotide repetitive motif in the human genome. Mutational targets were constructed containing TC/AG alleles up to 20 units in-frame within the coding region of the herpes simplex virus thymidine kinase (HSV-\(tk\)) gene. These targets were incorporated into oriP shuttle vectors, which replicate episomally in human lymphoblastoid cells. The overall HSV-\(tk\) mutant frequencies measured after 10 population doublings in cells derived from a clinically normal donor were slightly increased over the background of mutations recovered in Escherichia coli. DNA sequence analyses revealed that replication of TC/AG vectors in human cells increased the mutation frequencies at the microsatellite motif up to 3-fold, relative to background. Additionally, the median HSV-\(tk\) mutation rate of single-cell clones carrying the [TC/AG]\(_{17}\) vector was significantly different from that of clones harboring the control vector. The median rate of allele length alterations within the [TC/AG]\(_{11}\) tract was 2 \(\times\) 10\(^{-3}\) mutations/cell generation, with an equivalent rate of deletion and expansion mutations. In contrast, a [GT/CA]\(_{10}\) vector showed no increase in microsatellite mutation frequency after replication in human cells, and mutation rates of clones carrying a [GT/CA]\(_{10}\) vector were not significantly different from controls. Intriguingly, replication in human cells of all microsatellite-containing vectors resulted in elevated mutation frequencies at the downstream HSV-\(tk\) coding sequence of up to 20-fold, an effect not observed for the control vector. These results demonstrate that the frequency of mutational events at TC/AG motifs is slightly greater than at GT/CA motifs of similar allele length. This is the first report on our knowledge of the mutation rates at TC/AG microsatellite alleles in eukaryotic or prokaryotic cells.

INTRODUCTION

Neoplastic progression is accompanied by the generation of mutations to increase genetic diversity, followed by the selection and clonal expansion of mutant subpopulations (1). Microsatellite DNA sequences, or STR\(^3\) sequences, have emerged as powerful tools for detection of the mutational processes that arise during neoplastic development (2). Chromosomal deletions specific to tumor cells can be analyzed by loss of heterozygosity studies using polymorphic microsatellite DNA markers (3). Alterations in allele lengths of microsatellites, or MSI, can be used to detect a genome-wide instability in tumor cells caused by defective DNA repair (4, 5). The ability to use microsatellite DNA markers combined with the sensitivity of PCR allows the early detection of loss of heterozygosity or MSI in tumor DNA present in patient blood and urine samples (6–9). Despite the rapid development of microsatellite-based diagnostic techniques, little is known of the basic molecular biology of this class of genomic sequence. Many microsatellite sequences in the human genome are located in 5\(') transcribed but untranslated regions and introns of genes (10). Such sequences can adopt alternative DNA forms and have been postulated to be involved in transcriptional regulation and/or chromosomal organization (10, 11). Therefore, repetitive sequences in the genome may have a functional role that can be disrupted during neoplastic progression.

Regardless of their precise genetic function, measurements of microsatellite mutation rates in normal somatic cells are required to fully exploit these sequences for clinical cancer diagnoses. The high degree of polymorphism of microsatellite sequences in the human population is suggestive of a relatively high rate of germ-line variation (12), but only limited information is available for mutation rates in somatic cells. To date, only GT/CA dinucleotide repeats have been examined quantitatively for genetic stability in normal (13) or nontumorigenic (14, 15) mammalian cell experimental systems. Among dinucleotide repeat sequences, the TC/AG and GT/CA motifs are the most abundant, constituting 0.2 and 0.5% of the human genome, respectively (16, 17). Moreover, the two types of dinucleotides are dispersed throughout the genome in a similar manner (18, 19). The relative abundance of TC/AG microsatellites in the human genome and their ability to form DNA structures distinct from GT/CA microsatellites (20) prompted us to perform a comparison of the somatic mutation rates of TC/AG and GT/CA dinucleotides in nontumorigenic human cells. Our strategy quantitated the mutability of the microsatellite sequences relative to coding sequences within the same genetic target, the HSV-\(tk\) gene. This bimodal target was incorporated into an oriP shuttle vector that replicates episomally in EBV-transformed lymphoblastoid cells. Episomal vectors avoid potential complications because of random vector integration near endogenous microsatellite sequences. All mutations within the HSV-\(tk\) coding region that inactivate the HSV-TK protein, including those deletions or expansions within the dinucleotide motifs that generate alternative reading frames, are selected in bacteria, allowing for characterization of mutants by direct DNA sequence analyses. Our results show that, in the nontumorigenic lymphoblastoid cell line studied, the mutation frequencies and rates at repetitive dinucleotide sequences up to 20 units in length are similar to those of the unique sequences at the HSV-\(tk\) locus. However, TC/AG sequences were more mutable than GT/CA sequences of similar allele lengths. Intriguingly, we observed that the simple presence of the microsatellite sequences altered the mutability of the downstream HSV-\(tk\) coding sequence in these episomally replicated vectors.

MATERIALS AND METHODS

Reagents. Oligonucleotides used to insert various STR sequences were synthesized by Integrated DNA Technologies, Inc. (Coralville, IA) and the Macromolecular Core Facility (Pennsylvania State University College of Medicine). All restriction endonucleases were supplied by Life Technologies, Inc. (Gaithersburg, MD) and used according to the manufacturer’s instructions. All DNA sequence analyses were performed manually by the plasmid dideoxy sequencing method using Sequenase 2.0, according to the manufacturer’s protocol (Amersham Life Science, Inc., Arlington Heights, IL). FUdR, chloramphenicol, and hygromycin B were purchased from Sigma Chemical Co. (St. Louis, MO). Proteinase K and gentamicin were purchased from Life Technologies. Fetal bovine serum was purchased from Hyclone Laboratories Inc. (Logan, UT).
Cell Culture. LCL721, an EBV-transformed human lymphoblastoid cell line established from a clinically normal female donor (21), was cultured as described (22). Routine tests for Mycoplasm contamination were performed, and cultures were negative throughout this study. Balb 291, a murine fibroblast cell line in which the hygromycin phosphotransferase gene has been integrated into the genome (23), was cultured in DMEM supplemented with 10% fetal bovine serum, 50 μg/ml gentamicin, and 150 μg/ml hygromycin.

Construction of STR Shuttle Vectors. Vectors containing various STR sequences in the 5’ coding region of the HSV-tk gene were constructed by oligonucleotide site-directed mutagenesis (24) of plasmid pGTK4. pGTK4 is identical to the pGTK2 plasmid (25) except for a G to T base substitution at position 195 of the HSV-tk gene, which eliminates a potential internal start site for translation of the HSV-TK protein. The site-directed oligonucleotides were designed to insert a [GT/CA]9 or a [TC/AG] 9 in-frame between bases 111 and 112 of the HSV-tk gene (Fig. 1). The sequence of the HSV-tk sense strand only will be referred to from this point. These oligonucleotides also introduced a silent C→A substitution at position 117, generating an AvrI site at position 113 and allowing for detection of positive clones. A phagemid with an out-of-frame [GT]9 insert was also constructed as above. The HSV-tk gene of each construct was subcloned into the pND123 shuttle vector (26). The resultant pJY shuttle vectors are depicted in Fig. 1A. Two positive clones possessing the [GT]10 (pJY2 and pJY2A) and [TC] 9 (pJY4 and pJY4A) inserts were chosen to represent two independent DNA preparations of each STR vector. Vector pJY3, containing the [GT]10 insert, served as the positive control. Vector pJY1, constructed by subcloning the HSV-tk gene of pGTK4 into pND123, contained no STR insert and served as the negative control. Vectors pJY2.1 ([GT]10 insert) and pJY2.2 ([GT]10 insert) were derived from pJY2 as spontaneous STR frameshifts in Escherichia coli by sequential selection, first in the presence of 40 μg/ml trimethoprim, a folate analogue that selects for TK+ activity (27). Vectors pJY4.2 ([TC] 9 insert) and pJY4.3 ([TC] 9 insert) were derived from pJY4 in a similar manner. All vectors were propagated in recA13, upp, tdh E. coli strain FT334 (28). The sequence between HSV-tk positions 83 and 267 of each pJY vector was confirmed by DNA sequence analysis prior to use in the mutagenesis assay.

Manipulations of pJY Vectors in Lymphoblastoid Cells. pJY vectors (10 μg) were introduced into 107 log phase LCL721 lymphoblastoid cells by electroporation (22). After incubation for 48 h at 37°C, cells harboring plasmid were selected in medium containing hygromycin (300 μg/ml) for 5–7 days. After selection, an aliquot of the culture was removed for cloning, and the remaining cells were cultured in complete medium containing hygromycin (150 μg/ml) for a total of 10 population doublings. LCL721 cells bearing plasmid were cloned as single cells by two methods. The first method involved cloning in 0.35% agarose over a feeder layer of Balb 291 fibroblasts. Cloning efficiencies ranged from 0.1 to 0.8% under selective conditions (50 μg/ml hygromycin). The second method involved cloning cells in 96-well dishes by the limiting dilution method (29). Cloning efficiencies varied from 8 to 24% with selection (100 μg/ml hygromycin). To ensure that the clones originated from a single cell, we accepted only those cell densities that produced a fraction of negative wells >0.65 under nonselective conditions. Clones for mutational analyses were expanded from dishes seeded at identical densities in the presence of hygromycin. Clones from both the agarose and limiting dilution methods were expanded 28–35 generations in complete medium containing 150 μg/ml hygromycin.

pJY vector DNA was isolated from 2–3 × 108 lymphoblastoid cells by an alkaline extraction method (30) and incubation of the cell lysate with Proteinase K (100 μg/ml) for 2 h at 37°C. The nucleic acid pellet containing RNA, pJY vector DNA, EBV DNA, and mitochondrial DNA was resuspended in 10 mM Tris-HCl, 1 mM EDTA (pH 8.0) and digested with 25–35 μg RNase A and 150 units Xbal at 37°C for 2 h. Xbal cleaves both EBV and mitochondrial DNA. After ethanol precipitation, the lower molecular weight RNA was removed by centrifugation through a Microcon-100 ultrafiltration device (Amicon, Inc., Beverly, MA) to a final volume of 25–50 μl.

HSV-tk Mutational Analyses of pJY Vectors. To score for preexisting mutations created by the propagation and selection of our vectors in E. coli, background HSV-tk mutant frequencies were determined for each pJY construct as described previously (25). To determine HSV-tk mutant frequencies

---

**Fig. 1. pJY shuttle vectors and representative DNA sequence analyses.** A, pJY shuttle vectors. The positions of the bacterial (ColE1) and EBV (oriP, unfilled box) origins of replication are indicated. The chloramphenicol acetyltransferase gene (cat, hatched box) allows for selection in bacteria. The hygromycin phosphotransferase gene (hph, cross-hatched box) allows for selection in human cells. The filled box represents the coding region of the HSV-tk gene, and the gray box represents the 3′ noncoding region. The region from the EcoRI site at position 9542 to the AvrI site at 1165 is enlarged to emphasize important domains of the HSV-tk gene. STR sequences were inserted in-frame between positions 111 and 112. Single underline, an extra GT repeat unit; double underline, two extra TC repeat units flanking the STR sequences. +, AvrI site introduced during site-directed mutagenesis, which allows for identification of STR-containing clones. G, BglII; A, Avul; B, BssHII. B, autoradiogram showing representative pJY2 DNA sequence analyses. First panel, loss of 1 repeat unit ([GT]9); second panel, gain of four repeat units ([GT]14); middle panel, no change within the STR locus.
of pIV vector DNA isolated from lymphoblastoid cells, 1.5–2 μl of purified DNA was used to transform FT334 by electroporation and plated on VBA selective media. Plating by this method uses the presence of FUDR to select for bacteria with a plasmid-derived HSV-TK phenotype. Detectable mutations include deletions or expansions of any number of repeat units within the STR motifs that are not a multiple of three, as well as base substitutions, frameshifts, deletions, and rearrangements within the 1350-bp HSV-tk gene and promoter sequence (26–28). The HSV-tk mutant frequency is defined as the number of FUDR-resistant and chloramphenicol-resistant colonies divided by the total number of chloramphenicol-resistant colonies. For LCL721 clones, the mutation rate was calculated as the mutant frequency divided by the number of generations undergone by that clone at the time of pIV vector isolation. Differences in mutation rates among clones carrying different vectors were analyzed statistically using the nonparametric Mann-Whitney test. Both background pIV vector DNA and pIV vector DNA obtained after replication in lymphoblastoid cells were used to isolate independent mutants for mutational spectra as described previously (25). FUDR<sup>r</sup> mutants with large rearrangements (greater than ~30 bp in size) of the HSV-tk gene were detected by digestion of the mutant pIV vector DNA with BglII and Avai restriction endonucleases (Fig. 1A). DNA sequence analyses between HSV-tk positions 83 and 267 were performed on all remaining mutants (Fig. 1B). Differences in proportions of specific types of mutations between different vectors or between different hosts were analyzed statistically using Fisher’s exact test (two-tailed).

**RESULTS**

**Overall HSV-tk Mutant Frequencies for TC/AG and GT/CA Microsatellite-containing Vectors Replicated in LCL721 Cell Populations.** pIV shuttle vectors were constructed containing poly TC/AG and poly GT/CA alleles inserted in-frame or a poly GT/CA allele inserted out-of-frame (positive control) between positions 111 and 112 of the HSV-tk mutational target gene (Fig. 1A). These STR inserts are flanked 5' by an extra GT/CA unit and 3' by two extra TC/AG units. The total allele length and sequence of the HSV-tk sense strand for each vector will be reported from this point. These vectors, along with a negative control vector containing no inserted repeats, were propagated in *E. coli*. Background HSV-tk mutant frequencies representing mutations arising spontaneously in *E. coli* are presented in Table 1. DNA from the same vector preparations was also introduced into LCL721 human lymphoblastoid cells, and plasmid-bearing populations were expanded for 10 doublings. Shuttle vector DNA was isolated, and HSV-tk mutant frequencies were determined by selection in bacteria. These data are reported as the observed human frequencies of HSV-tk mutant frequencies in *E. coli* populations. Background HSV-tk mutant frequencies were determined by selection in bacteria. For the [TC]<sub>11</sub> vectors isolated from human lymphoblastoid cells, 1.5–2 fold above background. Corrected mutant frequencies of the [TC]<sub>11</sub> and the [GT]<sub>10</sub> vectors isolated from human cells increased only 2-fold above background (Table 1). Replication of the [GT]<sub>10</sub> vectors in human cells resulted in a 2-fold or less increase in the corrected mutant frequencies, relative to *E. coli* (Table 1). Overall, for all vectors examined, the HSV-tk mutant frequencies determined after replication in human cells were slightly elevated above background.

**STR Specificity Analyses of TC/AG and GT/CA Vectors in Lymphoblastoid Populations.** Mutation specificities of independent preparations of both the [TC]<sub>11</sub> and [GT]<sub>10</sub> vectors were generated by DNA sequence analyses (Fig. 1B) and are presented in Table 2. This rigorous approach permitted us to categorize mutants into either those with alterations in the STR motif or those with normal microsatellite allele lengths and alterations in the coding region of the HSV-tk target gene. DNA sequence analyses also allowed us to unequivocally determine the specificity of changes within the STR motifs. For the [TC]<sub>11</sub> vectors in *E. coli*, a partiality for deletions of one dinucleotide unit at the STR locus was detected (Table 2). We observed a significant difference in proportions of STR expansions relative to deletions between the two independent preparations of the [TC]<sub>11</sub> vectors (pIV4 and pIV4A) in *E. coli* (P = 0.04, Fisher’s exact test), indicating the stochastic nature of these microsatellites during propagation in *E. coli*. However, we observed no significant differences in the STR specificities of these combined [TC]<sub>11</sub> vectors when comparing mutants obtained from *E. coli* with those from human cells.

Mutants isolated from the [GT]<sub>10</sub> vectors propagated in *E. coli* showed preference for deletion mutations as opposed to expansion mutations at the STR motif with most having a deletion of one (GT) unit (Table 2). After replication in human cells, only four mutational events from each of the [GT]<sub>10</sub> vectors occurred within the STR locus (Table 2); therefore, the STR specificities for these individual vectors cannot be assessed statistically. However, combining the STR mutational data for both [GT]<sub>10</sub> vectors reveals that there are no significant differences between human and *E. coli* cells when comparing deletions relative to expansions.

**STR versus Coding Region Analyses of TC/AG and GT/CA Vectors in Lymphoblastoid Populations.** The proportion of mutants with changes in either the STR or coding region (Table 2) was multiplied by the corrected frequencies (Table 1) to determine region mutation frequencies for each vector. In *E. coli*, mutation frequencies at dinucleotide STR sequences of 10–11 units were ~10-fold greater than frequencies of mutation in the HSV-tk coding sequences (Table 3). As the TC allele length increased, the STR:coding ratios increased to ~30-fold in *E. coli*. In contrast, after replication in human cells,
such large biases for mutations occurring at the STR locus were not observed (Table 3).

The mean STR mutation frequency measured after replication in human cells for 10 population doublings for the two [TC]$_{11}$ vector preparations ($2.2 \times 10^{-5}$) was 3-fold greater than background (Table 3). The STR mutation frequency of the vector containing the longer [TC]$_{30}$ allele was $1.7 \times 10^{-3}$, only 2-fold above background. In contrast, the mutation frequencies measured at the HSV-tk coding sequence were increased $\sim 10$–20-fold after replication of all [TC] vectors in human cells, relative to background (Table 3). The combined mutational specificities of both [TC]$_{11}$ vectors (Table 2) showed a significant difference between E. coli cells and human cells when comparing the proportions of mutations within the microsatellite locus relative to the coding locus ($P = 0.0004$, Fisher’s exact test). Specifically, there were proportionally fewer mutational events within the STR region in human cells (93%) for the E. coli preparations ($2.2 \times 10^{-5}$) after replication in human cells, a bias not observed in E. coli (Table 3).

For the [GT]$_{10}$ vectors, no increased mutation frequencies at the STR motif, relative to background, were measured after replication of these vectors in human cells for 10 population doublings (Table 3). Again, it is intriguing to note that the coding region mutation frequencies of these vectors isolated from human cells were $\sim 10$–20-fold greater than the E. coli coding region mutation frequencies. Combining the mutational specificities of both [GT]$_{10}$ vectors (Table 2) revealed that the proportion of mutations occurring at the STR locus relative to the coding locus was significantly different between E. coli and human cells ($P < 0.0001$, Fisher’s exact test). As was the case for the [TC]$_{11}$ vectors, there were proportionally fewer mutations occurring at the STR locus in human cells (24%) than in E. coli cells (90%), and the coding region mutation frequencies were elevated above the control only after replication of the microsatellite vector in human cells (Table 3).

We also compared differences between [TC] and [GT] vectors of similar allele lengths. After replication in human cells, there was an extremely significant difference in the proportion of mutations occurring at the STR region relative to the coding region between the [GT]$_{10}$ and [TC]$_{11}$ vectors ($P = 0.0002$, Fisher’s exact test). More mutations occurred at the [TC]$_{11}$ STR locus (68%) than at the [GT]$_{10}$ STR locus (24%) in human cells, a difference not observed in E. coli.

### Analyses of the Stability of TC/AG and GT/CA Microsatellites in Clonal Lymphoblastoid Cell Populations.

To quantitate mutation rates at the episomal vector loci, plasmid-bearing lymphoblastoid cells were cloned as single cells and expanded for a total of 28–35 generations. The median mutation rate (mutation frequency/cell generation) of clones carrying the [TC]$_{11}$ vector ($10^{-6}$) was significantly different from that of the pND123 parent vector ($P = 0.01$, Mann-Whitney test; Fig. 2). However, the median mutation rates of the clones carrying [GT]$_{10}$ vectors did not vary $>2$-fold from that of those carrying the parent vector.

To determine the specificity of mutants arising during expansion of single-cell clones carrying the [TC]$_{11}$ vector, independent mutants from clones representing the lowest, the median, and the highest mutation rates were analyzed analogously to those of the parental populations. The greatest variation among the clones was in the mutation rate of large (>30 bp) HSV-tk deletions/insertions, which ranged from $0.24 \times 10^{-6}$ mutations/generation in the clone with the lowest mutation rate to $67 \times 10^{-6}$ mutations/generation in the clone with the highest mutation rate. In contrast, the coding region mutation rates varied $33$-fold (from 0.23 to $7.7 \times 10^{-6}$), and the STR region rates varied only $9$-fold (from 0.59 to $5.2 \times 10^{-6}$). The median STR mutation rate among the three clones was $2 \times 10^{-6}$ mutations/generation, compared with a median coding region mutation rate of $0.47 \times 10^{-6}$ mutations/generation. For the clones examined, the median rate of STR mutations ($0.91 \times 10^{-6}$) was similar to that of STR deletions ($0.69 \times 10^{-6}$).

### DISCUSSION

This is the first report of quantitative spontaneous somatic mutation rates at TC/AG microsatellite motifs. The median rate of allele length alterations within the [TC/AG]$_{11}$ tract was $2 \times 10^{-6}$ mutations/cell generation in human lymphoblastoid cells. The overall median HSV-tk mutation rate of clones carrying the [TC]$_{11}$ vector was significantly higher than that of clones harboring the parent vector (Fig. 2).
MICROSATELLITE MUTAGENESIS IN HUMAN CELLS

2. This result is consistent with the increased frequency of microsatellite changes with increasing allele length observed for GT/CA loci in yeast (31). Analyses of TC/AG vector-containing human cell populations revealed that the frequencies of mutations specifically at the TC/AG loci were 2–4-fold greater than those of the E. coli background (Table 3). Interestingly, the HSV-tk coding region mutation frequencies for all STR-containing vectors replicated in human cells were elevated 10–20-fold above those of the E. coli background (Table 3). Thus, in this episomal vector system, the simple presence of the microsatellite motifs increased the mutation frequency at downstream unique coding sequences in human cells, an effect not observed in E. coli.

Our studies demonstrate a differential mutability of TC/AG and GT/CA microsatellites in human cells. Perhaps the most compelling evidence suggesting that the TC motifs are less stable than the GT motifs is shown by combining data from both [TC]_{11} and both [GT]_{10} vectors (Table 2) and comparing the proportions of mutations in the STR versus coding regions. Such a comparison reveals an extremely significant (P = 0.0002, Fisher’s exact test) difference between these vectors after replication in human cells, with proportionally more mutations occurring in the STR motif for the [TC]_{11} vectors (68%) than for the [GT]_{10} vectors (24%). This difference was not observed between the TC/AG and GT/CA loci during propagation in E. coli. Moreover, clonal analyses of mutation rates revealed significantly different mutation rates for cells carrying the [TC]_{17} vector but not the [GT]_{16} vector, relative to the control cells (Fig. 2). Because both microsatellite sequences were inserted at the same position in the HSV-tk gene, the sequence context surrounding these STR loci is similar (Fig. 1), and therefore the observed instability differences are most likely attributable to the composition of the microsatellites.

The favored mechanism to explain alterations in microsatellite allele size is slipped strand mispairing between repeat units during replicative or repair DNA synthesis (32–34). The observed differential mutagenesis between TC/AG and GT/CA sequences may result from different probabilities of either utilization of misaligned intermediates by DNA polymerases or escape of the misaligned intermediate from DNA repair proteins (34). Although TC/AG and GT/CA motifs are of identical G-C content, poly TC/AG repeats can form triplex DNA, whereas poly GT/CA repeats can form Z-DNA (20). Although [TC/AG]_{12} motifs in plasmid DNA can form triplexes in E. coli (35), whether these secondary DNA structures are formed within our episomal vector in human cells remains to be tested. The TC/AG motifs also are asymmetric in that one DNA strand is composed of polypurine residues and the other of polypyrimidine residues. In contrast, GT/CA motifs are symmetric in that both strands are composed of alternating purine-pyrimidine residues. Any of these structural differences in DNA may increase the probability of misaligned intermediates forming in poly TC/AG tracts, relative to poly GT tracts, during DNA synthesis (36, 37). Alternatively, mutational intermediates formed in TC/AG microsatellites may be repaired less efficiently than those in GT/CA sequences. Although mismatch repair proteins have been shown to recognize and remove loops in GT sequences (31, 38–40), the ability of TC loops to serve as substrates has not been examined. However, mismatch repair studies in E. coli indicate that misaligned TC/AG and GT/CA intermediates are repaired similarly. Finally, formation of triplex DNA has been shown to cause an increased level of mutagenesis that is dependent upon a functional nucleotide excision repair pathway (41).

Other model systems have examined the mutation rate of GT/CA sequences in eukaryotic cells. In a wild-type yeast strain, an overall mutation rate of 1.6 × 10^{-5} events/cell division was reported for a vector containing a [GT]_{16} repeat (38, 39), a value similar to the median mutation rate of the [GT]_{16} vector in our system (5.2 × 10^{-6} mutations/cell generation; Fig. 2). Also in yeast, 73% of mutants that were obtained from replication of a vector containing a [GT]_{7.5} sequence had no alterations within the STR locus (31). We obtained similar results for the [GT]_{10} vector replicated in human cells in which 76% of mutants had no changes in STR allele lengths (Table 2). Mutagenesis at GT/CA sequences has been analyzed previously in normal human fibroblasts (13). In that system, the mutation rate at a [GT]_{17} sequence in an integrated vector was < 3.1 to 45 × 10^{-8} mutants/cell/generation. Because this system detected only one of three STR frameshifts, an upper limit was estimated to be as high as 2.7 × 10^{-6} mutants/cell/generation. In our analyses, we measured HSV-tk mutation rates ranging from 2.8 to 12 × 10^{-6} mutations/cell generation for a [GT]_{16} episomal vector (Fig. 2). However, as demonstrated by our DNA sequence analyses of lymphoblastoid cell populations bearing the [GT]_{10} vectors, a large proportion of inactivating mutations are HSV-tk rearrangements and point mutations and thus do not involve STR allele length changes. In fact, detailed sequence analyses of lymphoblastoid cell clones carrying the [TC]_{11} vector revealed that the mutation rate specifically at the STR locus ranged from only 0.59 to 5.2 × 10^{-6} mutations/cell generation, whereas the rate of all HSV-tk mutations ranged from 1.3 to 80 × 10^{-6}. Despite the differences in mutational detection, cell type, and vector construction between this and the previous study, our results corroborate the view that the inherent mutation rate of STR sequences < 20 units in length is low (< 10^{-6}) in nontumorigenic human cells. Furthermore, for the [TC]_{11} clones in our system, the median mutation rate at the STR locus was 2 × 10^{-6} mutations/cell generation, quite similar to the median mutation rate at the HSV-tk coding locus (0.47 × 10^{-6}).

In conclusion, we describe a facile system to measure quantitative...
mutation frequencies at microsatellites of varying DNA sequence. Our system has been designed to compare mutagenesis at various microsatellite loci to that at the coding region within the same target gene. Our mutational target has been incorporated into an EBV-derived oriP shuttle vector that is replicated episomally in EBV-transformed or EBNA-1-expressing human cell lines (42), thus eliminating variability attributable to random vector integration into the genome and allowing direct comparison of multiple cell lines. We have deduced somatic mutation frequencies and specificities at the two most abundant dinucleotide microsatellite loci. The utility of any microsatellite marker for molecular analyses of genetic alterations in tumor cells is ultimately dependent upon its inherent mutation rate in normal cells, for this determines the degree of marker polymorphism in the human population as well as the baseline frequency of alterations in allelic length. The intrinsic low rates of somatic mutation at microsatellites that we and others have measured can be used to estimate the probability of false-positive results in clinical samples that contain a vast excess of normal cell DNA over tumor DNA, thus allowing more precise determinations of test specificity (2). Additionally, our system can be used to clarify issues related to the distinction of low frequency MSI tumors (those that show MSI at one of five microsatellite loci examined) versus microsatellite-stable tumors (those that show MSI at zero of five loci) at various types of microsatellite motifs (4, 43). The demonstration that TC/AG loci are equally or slightly more mutable than GT/CA loci can be extrapolated to mean that TC/AG markers will be as informative as GT/CA markers in genomic studies. On the basis of the relative abundance of these two dinucleotide sequences (16, 17), the inclusion of TC/AG loci in human genomics may increase the number of available markers by ~40%.

ACKNOWLEDGMENTS

We thank Erin Gestl, Jill Hite, and Patricia Opresko for critical reading of the manuscript. We gratefully acknowledge the generous contributions made to the Jake Gitten Cancer Research Institute.

REFERENCES

Somatic Mutation Rates and Specificities at TC/AG and GT/CA Microsatellite Sequences in Nontumorigenic Human Lymphoblastoid Cells

Suzanne E. Hile, Guang Yan and Kristin A. Eckert


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/60/6/1698

Cited articles
This article cites 42 articles, 20 of which you can access for free at:
http://cancerres.aacrjournals.org/content/60/6/1698.full#ref-list-1

Citing articles
This article has been cited by 11 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/60/6/1698.full#related-urls

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

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

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.