Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 3354–3358, May 1987
Genetics

Sequence analysis of spontaneous mutations in a shuttle vector gene integrated into mammalian chromosomal DNA

(mutational hot spot/deletion/base substitution/retroviral vector)

CHARLES R. ASHMAN* AND RICHARD L. DAVIDSON†

Center for Genetics, University of Illinois College of Medicine, 808 South Wood Street, Chicago, IL 60612

Communicated by Theodore T. Puck, January 27, 1987

ABSTRACT **We have studied the molecular mechanisms of spontaneous mutations in mouse cells carrying a selectable bacterial gene. The mouse cells carry the *Escherichia coli* xanthine (guanine) phosphoribosyltransferase (*gpt*) gene in a retroviral shuttle vector integrated into chromosomal DNA in a proviral form. Cells with spontaneous mutations in the *gpt* gene were selected as resistant to 6-thioguanine and then were fused with COS cells for recovery of the mutant genes. Out of a total of 77 independent 6-thioguanine-resistant cell lines isolated in this study, vector sequences could be rescued from 43 of the mutant lines, and the base sequences were determined for the *gpt* genes in all 43 of these lines. There was a variety of mutational events among the mutant *gpt* genes sequenced. The most frequent mutational event was a deletion (in 29 of the 43 mutant genes), and the next most frequent event was a base substitution mutation (in 11 of the 43 mutant genes). Among the deletion mutants, the great majority represent deletions of less than 10 base pairs. In fact, 19 of the 29 deletion mutants had deletions of 3 base pairs, and among the mutants with 3-base-pair deletions, there was a very strong deletion hot spot appearing in 16 independent mutants. All 19 of the 3-base-pair deletions resulted in the "in frame" loss of an aspartic acid codon. Among the base substitution mutations, transitions and transversions occurred with approximately equal frequency. Our results raise the possibility that small deletions represent the predominant mechanisms by which spontaneous mutations occur in mammalian cells.**

Recently, several systems have been described that make possible the rapid and unambiguous determination of base sequence changes in genes that have undergone mutation in mammalian cells (1–5). In these systems, a target gene is introduced into the mammalian cells as part of a shuttle vector, which is capable of replication in both mammalian cells and bacteria. Following mutagenesis in the mammalian cell environment, the shuttle vector sequences are recovered from the mammalian cells and are introduced into bacteria, where large amounts of the mutant gene can be produced for sequence analysis. The major advantage of these shuttle vector systems is that large numbers of mutant genes can be rapidly isolated in a form that is convenient for DNA sequence analysis.

We recently described the development of a system of this type, which utilizes a retroviral shuttle vector (4). As a target gene, this vector contains the *Escherichia coli gpt* gene, which codes for the enzyme xanthine (guanine) phosphoribosyltransferase (XPRTase). The vector was introduced by infection into the mouse A9 cell line (6), which is deficient in hypoxanthine (guanine) phosphoribosyltransferase. A stable cell line (A9I2), which contains a single copy of the vector integrated into chromosomal DNA in a proviral form, was isolated. Mutants resistant to 6-thioguanine (Sgu) were induced in A9I2 following treatment of the cells with either ethyl methanesulfonate or BrdUrd. Mutant *gpt* genes were recovered from the 6-thioguanine-resistant (Sgu^r) cell lines by fusion with monkey COS cells (7). Following fusion, the simian virus 40 origin of replication within the retroviral vector is apparently activated, and the vector sequences are excised from chromosomal DNA by intrastrand recombination between the viral long terminal repeats (8). The resulting circular molecules have the properties of a bacterial plasmid, due to the presence of the pBR322 origin of replication and the *neo* gene within the vector. Thus, the plasmid molecules from Sgu^r cells could be recovered by transformation into *E. coli*. The DNA base sequences of a number of mutant *gpt* genes were determined directly from plasmid DNA by dideoxy extensions from a series of oligonucleotide primers. A major difference between this system and other shuttle vector systems currently being used to study mutagenesis is that the selectable gene is integrated into chromosomal DNA in mammalian cells rather than existing as an autonomously replicating molecule. In addition, selection for mutant target genes in this system takes place while the genes are still present in mammalian cells, rather than in bacteria.

In this study, we have isolated a series of Sgu^r A9I2 cell lines that arose spontaneously. Mutant *gpt* genes were recovered from a number of these cell lines, and the base sequence changes were determined by DNA sequencing. The predominant mutational event among the spontaneous mutants sequenced was a deletion rather than a base substitution mutation. The most striking feature of the spontaneous deletions observed was the existence of a very strong deletion hot spot. Other interesting features of the deletions are that they seem to be clustered, rather than distributed randomly throughout the *gpt* gene, and that most of the deletions are less than 10 base pairs (bp).

MATERIALS AND METHODS

Cell Culture. The basic cell culture medium was Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum. HAT medium (9) contains 0.1 mM hypoxanthine, 0.4 μM aminopterin, and 16 μM thymidine. G418 (10) medium contains 1 mg of G418 per ml. Selection for Sgu resistance was carried out in medium supplemented with 36 mM Sgu.

The construction of the A9I2 cell line has been described in detail (4). Briefly, this line was constructed by the introduction of the retroviral shuttle vector pZip-GptNeo into

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.

Abbreviations: HAT, hypoxanthine/aminopterin/thymidine; Kan^r, kanamycin-resistant; Sgu, 6-thioguanine; Sgu^r, 6-thioguanine-resistant; XPRTase, xanthine (guanine) phosphoribosyltransferase.

*Present address: Department of Radiation Oncology, University of Chicago Medical Center, 5841 South Maryland Avenue, Box 440, Chicago, IL 60637.

†To whom reprint requests should be addressed.

the mouse A9 cell line. The cells contain a single copy of the vector integrated into chromosomal DNA. Expression of the *gpt* gene and the *neo* gene from Tn5 enables the cells to grow in both HAT and G418 medium.

Mutant Isolation. A9I2 cells maintained in HAT medium were inoculated at a density of 10^5 cells per dish into 100-mm tissue culture dishes containing HAT medium supplemented with G418. After 10 days of growth in this medium (approximately five cell doublings), the cells from each culture were harvested individually and inoculated at a density of 10^5 cells per dish into a 100-mm dish containing G418 medium supplemented with 0.1 mM hypoxanthine/16 μM thymidine. After an additional 10 days of growth to allow for expression of mutations, the cells from each culture were harvested individually and inoculated at a density of 10^5 cells per dish into 100-mm dishes containing medium supplemented with 36 mM Sgu. After 10–14 days, Sgu^r colonies were isolated and expanded for further study. Only one Sgu^r colony was isolated from each set of dishes, to ensure the independent origin of the mutants analyzed.

Recovery and Analysis of Vector DNA. Vector sequences were recovered from Sgu^r A9I2 cell lines following fusion to COS-1 cells (11) as previously described (4). Two days after the fusion, low molecular weight DNA was extracted (12) from the fused cells and was used to transform *E. coli* strain DH-1 (13). Kanamycin-resistant (Kan^r) transformants (expressing the *neo* gene) were selected on plates containing that drug.

To determine the structure of the recovered plasmid molecules, small scale plasmid preparations were prepared from overnight cultures (2 ml) of the Kan^r transformants (14). Typically, six transformants were analyzed for each mutant. Restriction digest analysis of the recovered plasmids was carried out by using *Kpn* I digestion and agarose gel electrophoresis. Transformants appropriate for sequence analysis were identified as described below.

DNA Sequencing. Plasmid DNA was extracted from overnight cultures (15 ml) by the alkaline lysis method (14). After a brief RNase A treatment at room temperature, the preparations were extracted with phenol/chloroform and were concentrated by ethanol precipitation. Dideoxy sequencing reactions were carried out at 45°C on 1 μg of plasmid DNA using avian myeloblastosis virus reverse transcriptase. In this study, we utilized a series of four oligonucleotide primers located 5′ to and within the *gpt* coding sequence. A fifth primer located further upstream, about 150 bp 5′ to the coding region, was used to sequence deletion mutants that have a breakpoint 5′ to the coding region. Wedge-shaped (15) 8% acrylamide sequencing gels have been used to reliably resolve 150–200 bases.

RESULTS AND DISCUSSION

Mutant Isolation and Recovery. A total of 77 independent, spontaneous Sgu^r A9I2 cell lines were isolated in this study. In nine independent experiments, Sgu^r cells were observed at a frequency of approximately 2 mutants per 10^5 cells plated in Sgu (mean of 1.9 mutants per 10^5 cells, variance of 0.8). Following isolation of the mutants, attempts were made to rescue vector sequences from each mutant line by COS cell fusion. However, it was found that it was possible to rescue vector sequences from only 43 of the 77 mutants.

To obtain more information about the structure of the vector sequences in the mutant lines from which vector could not be rescued, the G418 sensitivity of each of the mutant cell lines was tested. It was found that 33 of the 34 mutants from which vector could not be recovered were no longer G418-resistant, whereas 42 of the 43 mutants from which vector was recovered had retained their G418 resistance. Thus, there was an almost perfect correlation between retention of G418 resistance and the ability to rescue vector sequences from these cell lines. It is thought that the mutants from which vector could not be recovered arose as a result of the total or partial deletion of vector sequences; this deletion could have resulted from either a deletion event or a loss of the chromosome into which the vector had integrated. High frequency deletion events involving retroviral sequences have been observed in other studies (16, 17).

To determine the alterations in the *gpt* genes in mutant cells from which vector sequences could be recovered, the *gpt* genes were sequenced. We have observed that when vector molecules are recovered from Sgu^r A9I2 cells about 50% of the recovered plasmids give a three-band *Kpn* I restriction pattern (3.5, 0.9, and 0.7 kilobases). This is the pattern expected if the vector sequences are excised from chromosomal DNA as the result of homologous recombination between the viral long terminal repeats. The remaining 50% of the plasmid molecules typically give a variety of *Kpn* I restriction patterns. When plasmid molecules having a three-band *Kpn* I pattern were observed, one of these plasmids was used for DNA sequence analysis.

It would be expected that cases would arise in which it would not be possible to rescue a plasmid having the expected structure because of deletions or base substitutions that have occurred in the mammalian cells. In this study we observed 2 cases (out of 43) in which Kan^r transformants were obtained from Sgu^r mutants, but none of the plasmid molecules gave the appropriate three-band *Kpn* I restriction pattern. In both cases, however, it was observed that about 50% of the recovered plasmids had the same restriction pattern whereas the remaining plasmids had a variety of patterns. In these cases, two isolates of each of the majority-type plasmids were sequenced, and it was found in each case that the plasmid had a deletion that would be expected to alter the *Kpn* I restriction pattern. Because of the high frequency with which these deletions were represented among the plasmid molecules recovered from the two mutants, it was assumed that these deletions arose in the mammalian cell prior to the recovery process.

The base sequences of the *gpt* genes have been determined for all 43 of the mutant cell lines from which vector sequences were recovered. A variety of mutational events were observed among the mutant *gpt* genes. The most common mutational event was a deletion, and this was observed in 29 of the 43 mutants. The next most common event was a single base substitution, and this was observed in 11 of the 43 mutants. In addition to these events, 1 mutant was observed to have an insertion of a single base, while 2 mutants had multiple base changes.

Spontaneous Deletions. Information about the location and size of the deletion mutations is presented in Table 1. One interesting feature of the deletions is that the great majority of them are relatively small. It should be possible to rescue vector sequences from Sgu^r A9I2 cell mutants that have deletions of the entire *gpt* gene (456 bp) (18–20) plus about 800 bp of flanking sequences on each side. However, of the 13 different classes of spontaneous deletions that we have observed (representing 26 of the 29 deletion mutants), 10 are less than 10 bp in length. To date, the largest deletion mutant from which vector sequences have been recovered has a deletion of 1256 bp. (Very large deletions affecting not only the *gpt* gene but also genes necessary for the rescue of vector sequences in bacteria would not be recovered in this system.)

In addition to their clustering in terms of size, the deletions appear to be nonrandomly distributed in the *gpt* gene. As can be seen in Table 1, there are four regions in the gene where 2–3 different deletions either share breakpoints or have breakpoints within 15 bases of one another. It appears that there are regions within the gene where deletions are more

Table 1. Size and location of spontaneous deletions

Isolates, no.	Bases deleted, no.	Base deleted	
		5′ most	3′ most
1	1256	−38, −37, −36	1218–1220
1	1	24	24
1	2	107	108
1	7	107	113
1	7	120–123	126–129
1	7	149, 150	155, 156
1	8	150, 151	157, 158
1	17	210–212	226–228
1	8	213, 214	220, 221
1	59	228, 229	286, 287
16	3	366–369	368–371
1	3	370	372
2	3	381, 382	383, 384

The positions of the most 5′ base and the most 3′ base deleted in the *gpt* sequence are indicated. The first base of the *gpt* coding sequence is designated as base number 1. In several cases, the exact location of the deletion breakpoint could not be determined due to the presence of a repeated sequence at the breakpoints.

likely to occur, although the breakpoints of these deletions may not be exactly the same in each event.

The most striking feature of the deletion mutations is the presence of a strong deletion hot spot, accounting for 16 of the 29 mutations. This mutation was observed repeatedly even though care was taken to ensure that each mutant was an independent isolate. Independent cultures were established from a stock culture, which had been maintained in HAT medium. Following their establishment, each culture was grown in HAT medium for an additional 10 days. After this time, each culture was then grown for 10 days in the absence of HAT selection medium prior to the selection of Sgu^r mutants, and only one mutant colony from each independent set of cultures was analyzed. A possible explanation for the isolation of many mutants having the same change in base sequence would be that this mutation enables the cell to grow in both HAT medium and Sgu. This, however, is clearly not the case, as cell lines having this mutation are killed rapidly in HAT medium.

The base sequence and the possible sites of the breakpoints of the deletion hot spot are presented in Fig. 1, deletion a. The mutations occur in the sequence 5′-TGATGA-3′ and result in the deletion of one TGA (read in frame as the deletion of a GAT codon). The structure of the hot spot (with small tandem repeats) is similar to that of a very strong spontaneous mutation hot spot in the *E. coli lac I* gene (21). The *lac I* hot spot occurs at the tandemly repeating sequence 5′-CTGGCTGGCTGG-3′, and the mutations consist of either the addition or the deletion of one unit of four nucleotides (CTGG). The *gpt* hot spot differs from the *lac I* hot spot in the length of the repeat unit (three vs. four nucleotides), in the number of repeats (two vs. three), and in that only deletions have been observed at the *gpt* site. (It is conceivable that the addition of a TGA triplet at the *gpt* hotspot would not produce the mutant phenotype and would be undetected.)

Other deletions have been observed in the immediate vicinity of the *gpt* hotspot, and the breakpoints of these other deletions also are presented in Fig. 1. One of these, deletion b, removes the codon immediately adjacent to that deleted by the hot spot mutations and in fact extends into the range of the possible hot spot deletions. Deletion c (represented in two independent mutants) occurs at a site about 15 bp away. This deletion occurs in the sequence 5′-GTTGAT-3′, which is identical to the sequence in which the 5′ breakpoint of the hot spot deletions occurs. In fact, in both this and the hot spot deletion, the codon GAT is removed.

Since multiple deletions have been observed in the hot spot region, it seems likely that sequences other than the 5′-TGATGA-3′ sequence are involved in the generation of these deletions. This sequence is also present at another site in the gene (positions 261–266), and no deletions have been observed in that region. This also suggests the possibility that sequences outside the immediate vicinity of the hot spot play a role in the generation of deletions in that region. (An alternative possibility is that deletion of a TGA triplet at positions 261–266 occurs but does not produce the mutant phenotype and thereby remains undetected.)

Another interesting feature of the deletions in the hot spot region is that each of the three different classes of 3-bp deletions results in the loss of a single aspartic acid residue from the XPRTase protein. Each of these deletions is in frame and does not result in any other alterations in the amino acid sequence. The aspartic acid residues in this region may be essential for XPRTase activity. The fact that this region is a hot spot for spontaneous mutations is probably a consequence of both the particular region of the protein involved and some unique aspect of the DNA base sequence in that area.

One structural feature of the hot spot is reflected in the ambiguity in the definition of the ends of the deletion. As shown in deletion a of Fig. 1, the exact same in frame deletion of a GAT codon can be generated by deletions beginning at any one of four different positions (366, 367, 368, or 369). The fact that 3-bp deletions beginning at any of the four sites lead to the same alteration could contribute to the high frequency at which this deletion occurs. The results presented in Table 1 show that such ambiguity in defining the precise ends of a deletion also occurs in over 75% of the other deletions that we have observed in the *gpt* gene. Overall, these ambiguities are the consequence of the presence of short direct repeats at the ends of the deletions. The presence of such repeats at the ends of deletions has been observed in other systems, and it has been suggested that they play a role in the generation of the deletions (22, 23).

Spontaneous Base Substitutions. The location and types of base changes in the spontaneous, single base substitution mutations are presented in Table 2. Of the 11 base-substitu-

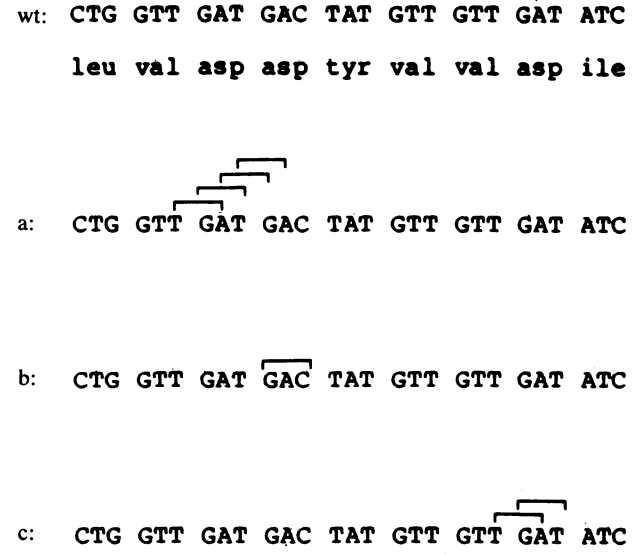

FIG. 1. Location of deletions in the spontaneous deletion hot spot region of the *gpt* gene. The DNA base sequence and amino acid assignments are given for positions 361–387 of the wild-type (wt) *gpt* gene. The possible locations of the three spontaneous deletions in this region are indicated by brackets above the sequences. Sixteen independent isolates of deletion a, one of deletion b, and two of deletion c have been sequenced.

Table 2. Location and base changes of spontaneous base substitution mutations

Base change	Position	Isolates, no.
G·C → A·T	92	1
	139	1
	242	1
A·T → G·C	119	1
	169	1
G·C → C·G	46	1
	139	1
	417	1
A·T → T·A	1	1
	52	1
A·T → C·G	257	1

For the designation of the numbering of the bases, see Table 1.

tion mutants sequenced, 6 are transversions and 5 are transitions. There was no obvious preference for any particular type of base substitution among the spontaneous point mutants sequenced. There was also no obvious site specificity among this class of mutations. The 11 mutations occurred at 10 different sites in the *gpt* gene. Two mutations, a G·C → C·G transversion and a G·C → A·T transition, were observed at base 139. The absence of any strong base or site specificity for spontaneous base substitution mutations differs from the situation in the *lac I* gene in *E. coli*, where there is a good correlation between the occurrence of 5-methylcytosine and the occurrence of G·C → A·T transition mutation hot spots (24). Our results also differ from those on spontaneous mutations of the *lac I* gene in an autonomously replicating shuttle vector, where base substitution mutations occurred exclusively at G·C pairs (25). In the present study, approximately half of the base substitution mutations occurred at A·T pairs.

Our results indicate that only a small fraction (approximately one-fourth) of the mutant *gpt* genes sequenced resulted from single base substitutions; the large majority resulted from deletions. The proportion of base substitution mutations observed in this study may be an underestimate because some base substitutions would not lead to an altered amino acid or would produce an amino acid substitution that permits enzyme activity. At present, the number of nonsense and missense mutants that we have isolated is not large enough to permit an accurate assessment of the degree to which base substitutions may be underestimated as the result of silent mutations in the *gpt* gene. It may be noted that a study of the base changes in a human tubulin pseudogene, a presumably neutral sequence, revealed a preponderance of base substitutions rather than deletions (26).

Other Spontaneous Mutations. In addition to the deletion and single base substitution mutants, there were three other spontaneous *gpt* mutants observed. One of these mutants was the result of the insertion of a single A·T base pair. This insertion occurred between bases 7 and 13, within a run of five A·T base pairs. Two mutants were found to have multiple base changes. One of these occurred at bases 49–54 and resulted in the change of the sequence 5′-CGTAAA-3′ to 5′-TTTAA-3′. The second mutation with multiple changes had five separate single base substitutions at sites throughout the *gpt* gene. Three of these base substitutions were G·C → A·T transitions, while the other two were A·T → C·G and G·C → T·A transversions. Two isolates of this mutant were sequenced, and both were found to have the same five base substitutions.

The pZip-GptNeo System. We have shown previously that mutations arise spontaneously at very high frequencies when shuttle vector sequences are rescued from mammalian chromosomal DNA following COS cell fusion (4, 27). Therefore, a point of concern during the development of this system was that mutations would be generated during the recovery process and that it would not be possible to distinguish these mutations from mutations generated in the A9I2 cells prior to recovery. Several lines of evidence strongly suggest that this is not a serious problem. In a previous study (4), we recovered vector sequences from Gpt^+ A9I2 cells and observed that about 50% of the plasmid molecules had intact *gpt* genes as determined by restriction digest analysis. The remaining 50% of the plasmids had gross rearrangements in structure, and many different plasmid structures were observed. It was found that all of the plasmids with an intact *gpt* gene produced a Gpt^+ phenotype in *E. coli*, whereas all of the plasmids without an intact gene were Gpt^-. Thus, it appears that most or all of the mutations generated during the recovery process are gross rearrangements and that many different rearrangements are produced. Those plasmid molecules that do not undergo gross rearrangements during the recovery process apparently do not suffer any damage that affects the ability of the *gpt* gene to function in *E. coli*. Therefore, our strategy in subsequent studies has been to sequence the "majority" plasmid recovered from the mutant cell lines on the assumption that this plasmid contains vector sequences that are identical to those in the mammalian cells. Examination of the sequences of the mutant genes recovered from Sgu^r cells suggests that this assumption is correct. If mutations were being generated during plasmid recovery, one would expect to observe mutants with base sequence alterations in at least two different sites. One mutation would have been generated prior to plasmid recovery and would have been responsible for the development of Sgu resistance, while the other mutation(s) would have been generated during recovery. However, of 91 mutant genes that have been recovered and sequenced to date (in this and other studies in our laboratory), only 1 has exhibited base sequence alterations at more than one site. Two isolates of this exceptional mutant were sequenced, and both were found to have the same base alterations. Thus, it is likely that even those alterations were generated in the mammalian cell prior to plasmid recovery.

A second line of evidence suggesting that mutations generated during the recovery process have not made a significant contribution to the mutations we have observed is that the mutational spectrum of *gpt* genes recovered from spontaneous mutants is very different from that of *gpt* genes recovered from ethyl methanesulfonate- and BrdUrd-induced mutants (C.R.A. and R.L.D., unpublished results). Unlike the spontaneous mutations, which are largely deletions, about 90% of the ethyl methanesulfonate- and BrdUrd-induced mutations are base substitutions. Since the same recovery procedure has been used in all the studies, this suggests that most or all of the mutations that we have characterized were generated in the mammalian cells prior to recovery of the vector.

In the context of the above discussion concerning generation of mutations during plasmid recovery, it is important to recall the characteristics of the integrated shuttle vector system that we are using. In this system, the mutant *gpt* gene with alterations leading to functional deficiency are "purified" by selection for Sgu resistance prior to recovery of the plasmid molecules. Since we have demonstrated that A9I2 cells contain only a single copy of the retroviral vector (4), each clone of Sgu^r cells isolated should have only one type of mutant gene, and any alterations generated during plasmid recovery would be against the background of a homogeneous population of that mutant gene. This is in marked contrast to the situation with shuttle vectors that replicate autonomously in mammalian cells. In such cases, the shuttle vector population in mammalian cells is at all times a mixed population, and vector molecules are not purified until they are recovered in bacterial transformants. Thus, identification of those genes

that had mutated in the mammalian cells prior to vector recovery, in contrast to genes altered during recovery, could be much more difficult with autonomously replicating vectors than with chromosomally integrated shuttle vectors.

In contrast to other studies on the molecular basis of spontaneous mutations in mammalian cells (for review, see ref. 28), the results of the present study indicate that, among the mutants from which vector sequences could be recovered, a high proportion (indeed the large majority) of spontaneous mutants in chromosomally integrated shuttle vector genes are the result of small deletions. It is possible that the difference in the results obtained reflects the different types of systems used. In general, the other systems used to study spontaneous mutations have involved either autonomously replicating shuttle vectors or endogenous chromosomal genes. In the first case, analysis of mutant shuttle vector molecules generally has been by means of procedures that identify only base substitution (nonsense) mutants and, therefore, exclude all deletion mutants (25). In the second case, analysis of mutant chromosomal genes generally has been based on Southern blotting procedures incapable of detecting small deletions (29). Thus, it is possible that small deletion events would have escaped detection or identification in the other types of systems. In contrast, the integrated shuttle vector system that we are using has no inherent bias (other than the requirement for the mutant phenotype) in the type of mutation that can be isolated and characterized. The results of the present study, therefore, raise the possibility that small deletions represent the predominant mechanism for generating spontaneous mutations in mammalian cells. The applicability of this conclusion to endogenous chromosomal genes remains to be tested in systems capable of detecting small deletions in such genes.

We thank Pamela Broeker for expert technical assistance. This work was supported by Public Health Service Grant CA 31781 from the National Institutes of Health.

1. Lebkowski, J. S., Clancy, S., Miller, J. H. & Calos, M. P. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 8606–8610.
2. Hauser, J., Seidman, M. M., Sidur, K. & Dixon, K. (1986) *Mol. Cell. Biol.* **6,** 277–285.
3. Glazer, P. M., Sarkar, S. N. & Summers, W. C. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 1041–1044.
4. Ashman, C. R., Jagadeeswaran, P. & Davidson, R. L. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 3356–3360.
5. Drinkwater, N. & Klinedinst, D. K. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 3402–3406.
6. Littlefield, J. W. (1964) *Nature (London)* **203,** 1142–1144.
7. Gluzman, Y. (1981) *Cell* **23,** 175–182.
8. Botchan, M., Topp, W. & Sambrook, J. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **44,** 709–719.
9. Littlefield, J. W. (1964) *Science* **145,** 709–710.
10. Davies, J. & Jiminez, A. (1980) *Am. J. Trop. Med. Hyg.* **29,** 1089–1092.
11. Davidson, R. L. & Gerald, P. S. (1976) *Somatic Cell Genet.* **2,** 165–176.
12. Hirt, B. (1967) *J. Mol. Biol.* **26,** 365–369.
13. Hanahan, D. (1983) *J. Mol. Biol.* **166,** 557–580.
14. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7,** 1513–1523.
15. Chen, E. Y. & Seeburg, P. H. (1985) *DNA* **4,** 165–170.
16. Shtivelman, E., Zakut, R. & Canaani, E. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 294–298.
17. Jolly, D. J., Willis, R. C. & Friedmann, T. (1986) *Mol. Cell. Biol.* **6,** 1141–1147.
18. Richardson, K. K., Fostel, J. & Skopek, T. R. (1983) *Nucleic Acids Res.* **11,** 8809–8816.
19. Pratt, D. & Subramani, S. (1983) *Nucleic Acids Res.* **1,** 8817–8823.
20. Jagadeeswaran, P., Ashman, C. R., Roberts, S. & Langenberg, J. (1984) *Gene* **31,** 309–313.
21. Farabaugh, P. J., Schmeissner, V., Hofer, M. & Miller, J. H. (1978) *J. Mol. Biol.* **126,** 847–863.
22. Streisinger, G., Okada, Y., Emrich, J., Tsugita, A., Terzaghi, E. & Inouye, M. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31,** 77–84.
23. Albertini, A., Hofer, M., Calos, M. P. & Miller, J. H. (1982) *Cell* **29,** 319–328.
24. Coulondre, C., Miller, J. H., Farabaugh, P. J. & Gilbert, W. (1978) *Nature (London)* **274,** 775–780.
25. Miller, J. H., Lebkowski, J. S., Greisen, K. S. & Calos, M. P. (1984) *EMBO J.* **3,** 3117–3121.
26. Hall, J. L., Dudley, L., Dobner, P. R., Lewis, S. A. & Cowan, N. J. (1983) *Mol. Cell. Biol.* **3,** 854–862.
27. Ashman, C. R. & Davidson, R. L. (1984) *Mol. Cell. Biol.* **4,** 2266–2272.
28. Thacker, J. (1985) *Mutat. Res.* **150,** 431–442.
29. Fuscoe, J. C., Fenwick, R. G., Ledbetter, D. H. & Caskey, C. T. (1983) *Mol. Cell. Biol.* **3,** 1086–1096.