Heteroduplex-induced mutagenesis in mammalian cells

Ursula Weiss and John H. Wilson

The Verna and Marrs McLean Department of Biochemistry, Baylor College of Medicine, Houston, TX 77030, USA

Received October 13, 1987; Accepted January 15, 1988

ABSTRACT

We have shown previously that heteroduplexes containing single-stranded loops are repaired efficiently in monkey cells, but not always correctly: 2% of the repair products acquired mutations within a 350 base-pair target (Weiss, U. and Wilson, J.H., *Proc. Natl. Acad. Sci. USA* 87:1123-1126, 1987). The structures of the mutant genomes, which are described here, are consistent with an error-prone repair system. The spectrum of mutations includes about 25% point mutations and 75% rearrangements, which consist of deletions, duplications, and substitutions. The mutations are clustered in the vicinity of single-stranded loops in the original heteroduplex. The high frequency of mutation, their clustering, and the positions of rearrangement endpoints suggest that the mutations were generated during repair of the heteroduplexes.

INTRODUCTION

Heteroduplex DNA is formed as a key intermediate in all described pathways of homologous recombination (1-4). Genetic and physical evidence indicate that mispaired and unpaired bases can be incorporated into heteroduplexes (5-8). Repair of the resulting abnormal structures is thought to contribute to classic genetic phenomena such as marker effects and gene conversion (9-12).

We have shown previously that SV40 heteroduplexes containing one or multiple single-stranded loops in the 350 base pair intron of the T-antigen gene undergo efficient repair in monkey kidney cells (13). More than 90% of the transfected heteroduplexes gave rise to plaques containing viral DNA of a single genotype. Although either strand could be used as the template for repair of the mismatch, there was nearly a 2:1 bias against the strand with the loop. Adjacent loops in multi-loop heteroduplexes were often co-repaired, consistent with the formation of excision-repair tracts extending up to 200-400 nucleotides in length. In this report, we examine the genomic structures of mutant SV40 virus that appeared among the progeny from transfected heteroduplexes. The structures of the mutant genomes are consistent with a

repair process that is initiated by endonucleolytic scission at or near the base of a single-stranded loop and that is completed aberrantly about 2% of the time.

MATERIALS AND METHODS

Cells, Viruses, DNAs

The construction of SV40 heteroduplexes, their transfection into CV1 monkey kidney cells and the analysis of progeny viral plaques have been described (13). In brief, heteroduplexes were formed from pairwise combinations of SV40 wild type and deletion mutant DNAs. DNA transfections were carried out with DEAE dextran and 0.01 to 0.05 nanograms of SV40 DNA per 60mm culture dish. Individual, well-isolated plaques were picked and then amplified by reinfection of CV1 cells in the presence of ^{32}P -orthophosphate. Labeled viral DNA was isolated according to Hirt (14), digested with restriction enzymes HindIII and TaqI, and subjected to electrophoresis on a 5% polyacrylamide gel.

DNA Sequencing

DNA samples were purified as described (15), and the double-stranded DNA was sequenced according to the method of Zagursky et al. (12) with modifications as described (15). Sequencing primers were 17-mers that hybridized within the coding region of the T antigen gene.

RESULTS

Twenty-two viral plaques, corresponding to 2% of the analyzed progeny from heteroduplex transfections, were identified as mutants by the abnormal migration of restriction fragments on polyacrylamide gels. As we have discussed before, this assay is quite sensitive to minor nucleotide changes (13,16). In this study, the assay readily separated two parental deletions that differ in length by two base pairs within a 400 base-pair fragment and permitted identification of single-base substitutions within a 700 base-pair fragment. As a further check on our ability to identify mutants in this way, we examined by sequence analysis 26 progeny virus with normal restriction patterns. These viruses, which were derived from 11 different input heteroduplexes, contained only the parental markers deduced from their restriction patterns: no additional mutations were found in 2000 base pairs corresponding to double-stranded DNA or in 525 base pairs corresponding to single-stranded DNA in the input heteroduplexes. These results confirm the sensitivity of the electrophoretic assay; however, the observed 2% mutation frequency must be considered

Table 1: Genotypic Classification of Mutants

Heteroduplex Type ^a	Total Plaques	Mutant Plaques ^b	Mutants Sequenced	Mutant Types			
				Point	Rearrangement ^c		
					Del	Dupl	Sub
Single	395	3	2	1	1	0	0
Double Cis	140	0	0	0	0	0	0
Double Trans	423	10	10	3 ^d	6	2	0
Triple	165	9	9	1	3	4	1
	<u>1123</u>	<u>22</u>	<u>21</u>	<u>5</u>	<u>10</u>	<u>6</u>	<u>1</u>

a) Heteroduplex type refers to the number and arrangement of single-stranded loops. "Single" indicates heteroduplexes containing one loop; "double cis" indicates heteroduplexes containing two loops on the same strand; "double trans" indicates heteroduplexes containing two loops on opposite strands; "triple" indicates heteroduplexes containing three loops, either in the trans/cis or the cis/trans configuration (13).

b) Mutant viral plaques were identified by the abnormal migration of their restricted DNA.

c) Del, deletion mutant; dupl, duplication mutant; sub, substitution mutant.

d) One point mutation was found in a duplication mutant (mutant I in Figure 1).

a minimum estimate, since some mutations undoubtedly escaped detection.

The sequence alterations in mutant genomes comprised single-base changes and a variety of rearrangements, including deletions, duplications, and a substitution (Table 1). The point mutations included 3 GC to AT transitions and two GC to CG transversions. Preferential mutagenesis of GC base pairs as opposed to AT base pairs was noted also when vector DNA was shuttled through mammalian cells (17,18). With the exception of one mutant (mutant L under Single Loop in Figure 1), all the rearrangements came from heteroduplexes with multiple loops and most rearrangements involved trans loops.

The rearrangements are illustrated schematically in Figure 1 to show their relationship to the structure of the initial heteroduplexes. The portions of each heteroduplex that remain in the mutants are indicated by thick lines and dashed lines; the rearrangement junctions lie at the right end of the thick line and the left end of the dashed line. Mutants A, B, C, D, F, G, K, L, and M are deletions, as indicated by the gap between the thick and dashed lines. Mutants H, I, J, N, O, and P are duplications, as indicated by the overlap between the thick and dashed lines. Mutant E is a substitution in which 17 nucleotides of unknown origin (see Figure 2) are present between the ends of the thick and dashed lines. The nucleotide sequences across the junctions are shown in Figure 2. Homologous nucleotides at the junctions are boxed; they range in number between 0 and 5, as is typical of nonhomologous (or illegitimate) junctions in mammalian cells (15).

Figure 1: Genomic structures of rearrangement mutants. The outline of the input heteroduplexes is shown. The portions of each heteroduplex that are retained are shown by the thick and dashed lines. The breakpoints that form the junction in each mutant lie at the right end of the thick lines and at the left end of the dashed lines. Straight arrows denote breakpoints within 15 nucleotides of the base of a loop; wavy arrows denote breakpoints that are separated from the nearest heteroduplex loop by more than 15 nucleotides; open arrows denote apparent endpoints near a loop in "complex" rearrangement mutants; dotted arrows refer to intervals containing a crossover. (One deletion mutant is not included in this figure because its endpoints were located too near the sequencing primers to allow their determination.)

Trans I, II, and III refer to the rearrangement mechanisms described in Figure 4. Trans I mutants contain deletions; mutant E has replaced the deleted sequence with 17 base pairs of unknown origin. Trans II and III mutants contain duplications. The Cis mutant contains a deletion with endpoints in the vicinity of two loops on the same heteroduplex strand. "Complex" mutants contain a crossover between two heteroduplex loops. Mutants in the category "other" have neither endpoint near the base of a loop.

Mutant L was derived from a heteroduplex containing a single loop 57 nucleotides long. Mutants A, F, G, J, and M were derived from a trans double loop heteroduplex in which the left loop (25 nucleotides) was separated from the right loop (57 nucleotides) by 196 base pairs. Mutants B and C were derived from a similar heteroduplex in which the identical loops were separated by 169 base pairs. Mutant I was derived from a heteroduplex in which the left loop (25 nucleotides) was separated by 38 base pairs from the right loop (27 nucleotides). Mutants D, E, H, K, O, P, and N were derived from two different triple-loop heteroduplexes that differed in the arrangement of the loops; however, in both heteroduplexes, the first loop (25 nucleotides) was separated by 38 base pairs from the second loop (27 nucleotides), which was separated by 131 base pairs from the third loop (57 nucleotides). The deletion mutants used here are described in detail elsewhere (13,24).

- A. 4885-TGCAATGTACTGCAAA¹Caatggcctgagt
gacataattggaCAAACTACCTACAGAGA-4652
- B. 4885-TGCAATGTACTGCAAA¹Caatggcctgagt
gacataattggaCAAACTACCTACAGAGA-4652
- C. 4877-ACTGCAACAAT¹ggcctgagtgtg
cataattggacaAACTACCTACAG-4655
- D. 4877-ACTGCAACAAT¹ggcctgagtgtgc
gtgacataattgGACAACTACCTA-4658
- E. 4877-ACTGCAACAAT¹ggcctgagtgtgc
gtgtgacataattGACAACTACCTA-4660
TACCTACAACAATGTA
- F. 4878-TACTGCAACAAT¹ggcctgagtgtg
cctacagagattTAAAGCTCTAAGG-4637
- G. 4922-CTGAGGTATTGCT¹tcttccttaaat
tggacaaactacCTACAGAGATTAA-4747
- H. 4819-GTGCTTACTGAGAT¹Gaatgcatgaaat
aatctggtgttGATGCAATGTACTGCA-4872
- I. 4824-TTGCTGTGCTTACT¹Gaggatgtagcatg
atgctcatcaacCTGACTTTGGAGGCTT-4934
- J. 4783-ATACAGGAAAG¹atccactgtgtg
gcaacaatggcCTGAGTGTGCAA-4851
- K. 4871-AACAATGGCTGAG¹gtgtgcaagaaaa
agaaaattatacAGGAAAGATCCACTT-4765
- L. 4635-AAATATAAAAT¹tttaagtgtata
gtaatgtgttaaCTACTGATTCT-4591
- M. 4751-GCTACTGCTTCGAT¹gtctttagaatg
gattgctttagaATGTGTTTGACT-4715
- N. 4769-CACTTGTGTGGT¹tgattgtactg
aaaatatgctcaTCAACCTGACTTT-4942
- O. 4813-CTGAGGATGAAG¹catgaaaaataga
gtgttgatgcaaTGTACTGCAAC-4869
- P. 4820-TGTGCTTACTGAG¹gatgaaatgcatg
tatttgcctcttCCTAAATCTG-4893

Figure 2. Nucleotide sequences across the junctions in the rearrangement mutants. In each case the top strand corresponds to the thick line in Figure 1 and the bottom strand corresponds to the dashed line. All sequences have their 5' end at the left. Sequences that are retained in the mutants are shown in upper case letters; sequences that are lost are shown in lower case letters. Homologous nucleotides at a junction are boxed. One nucleotide at the end of the top and bottom strands is numbered for reference to the SV40 nucleotide sequence (25,26).

The positions of the junctions relative to the loops in the input heteroduplexes appear to be nonrandom. Seven mutants--A, B, C, D, E, H, and L--have both endpoints (solid arrows) near the base of a loop; five mutants--F, G, I, J, and K--have one endpoint near the base of a loop and one endpoint (wavy arrow) more than 15 nucleotides away from the base of a loop; and two mutants (listed as "other") have neither endpoint near the base of a loop. Two additional mutants (listed as "complex") have their apparent endpoints (open arrows) near the bases of loops, but contain a "crossover" between the loops (dotted arrows), which complicates their interpretation.

The positions of the point mutations and the endpoints of the rearrangement mutations are illustrated in Figure 3A. In the figure, the positions of the single-stranded loops in the heteroduplexes are shown as thin lines; the 15 base pairs of flanking double-stranded DNA are shown as open boxes; arrows

Figure 3: Position of point mutations and rearrangement-mutation endpoints in relation to heteroduplex loops. A. Distribution of mutations relative to the single-stranded loops in the input heteroduplexes. The thin lines represent single-stranded DNA in heteroduplex loops; the open boxes represent 15 base pairs of flanking double-stranded DNA. The number of nucleotides in each loop are (from top to bottom) 25, 27 and 57. Arrows point to rearrangement endpoints (with the same code as in Figure 1). Filled circles indicate the positions of point mutations. B. Summary of distribution of rearrangement endpoints. The frequencies of apparent breakpoints relative to the base of a loop are summarized. The thin line represents 15 base pairs of single-stranded loop DNA, the open box represents 15 base pairs of flanking double-stranded DNA; the base of the loop coincides with the "0" position on the X axis. Each rearrangement endpoint is represented as a filled rectangle, except for the endpoints of the two "complex" mutants, which are represented by open rectangles.

indicate rearrangement endpoints (with the same code as in Figure 1); and filled circles show the positions of point mutations. Both classes of mutation appear to be clustered in the vicinity of the single-stranded loops in the input heteroduplex DNA. Four of the five point mutations are in a loop or at the base of a loop and most of the endpoints of the rearrangements are in or adjacent to a loop. The positions of the rearrangement endpoints are summarized in Figure 3B: 72% of the rearrangement endpoints lie within 13 nucleotides of the base of a loop, with a preponderance of endpoints at the

Figure 4: Proposed pathways for heteroduplex-induced rearrangements. In each heteroduplex the open boxes refer to flanking double-stranded DNA; the thin lines represent the left (L) and right (R) single-stranded loops; the closed box represents the interloop region. The arrows point to regions that incur a single-stranded break in the proposed mechanisms. Wavy vertical lines indicate the position of the junction. Each pathway generates a unique product: the Trans I pathway generates interloop deletions, whereas the Trans II, III, IV, and V pathways generate interloop duplications with characteristic structures.

The indicated pathways for rearrangement are illustrative only and are not unique. For example, in the Trans I pathway, the interloop region is shown as unwound, unpaired single strands. Unwinding in this pathway is not essential. The single strands from the loops could pair at their termini and prime repair synthesis, leading to a double-stranded bubble with the interloop region on one side and the L/R region on the other. Resolution of the bubble by removal of the interloop segment would generate the final structure. Similarly, the mechanism of end joining is left unspecified. End joining could occur by the pairing of terminal nucleotides of the single strands or, alternatively, by filling in the ends, followed by joining of the double-stranded ends (15).

exact base of the loop or within the adjacent single strand. If the distribution of endpoints were random, only 38% of the endpoints would be expected to lie so near the base of a loop.

DISCUSSION

The patterns of rearrangement in the mutants in this study are consistent with the operation of a repair enzyme that initiates repair by introducing a nick at or near the base of a single-stranded loop. Normally, the repair is completed accurately, but occasional mistakes are made: ends escape, find a partner, and join. The rearrangements can be understood as a consequence of the breakage of opposite strands, usually near the base of a loop, followed by joining of the broken ends. We have previously shown that end joining nearly always occurs within 15 nucleotides of the end (15). Five different ways of breaking opposite strands at trans loops along with the patterns of rearrangement resulting from end joining are shown in Figure 4. The existence of mutants in classes Trans I, II, and III (Figure 1) but not in classes Trans IV and V, may indicate that a step in the generation of rearrangements is polar. For example, if breakage always occurred on the 3' side of the mispaired region, breaks as shown in Trans IV and V would not occur. In this regard it is interesting to note that T4 Endonuclease VII nicks heteroduplexes with single-stranded loops immediately 3' of the mismatch in either the looped or the nonlooped strand (B. Kemper, personal communication). If breakage is polar, then defined heteroduplex (containing its Watson strand from one mutant and its Crick strand from another) should generate Trans I (deletion) or Trans II (duplication) mutants, but not both.

The scarcity of mutants at single loops and at cis loops may, in part, reflect the bias of loop repair, which is 2:1 in favor of excision of the loop. Generation of mutants from single loops and cis loops, according to the general scheme in Figure 4, would require that the nonlooped strand be broken.

Although the patterns of rearrangement are consistent with the aberrant operation of a mismatch-repair process, the possibility that mutagenesis occurred during transfection must be considered. The observed frequency of mutation and the distribution of endpoints argue against random mutagenesis during transfection. The mutation frequency observed in our experiments is 10- to 20-fold higher than that observed in transfection experiments with circular molecules containing similar-sized target regions (19,20). Although single-stranded regions may increase the sensitivity to transfection mutagenesis, random breakage of single-stranded regions would not account for the distribution of endpoints near the bases of loops, nor would it account for the average retention of 88% of the loop strand in those mutants in which the junction involves a loop. Whether any of the events critical to mutagenesis occurred outside the nucleus (as expected for transfection-associated muta-

genesis) could potentially be resolved by microinjection of heteroduplexes.

Formation of heteroduplexes seems to be relatively mutagenic in mammalian cells. Heteroduplexes with single-stranded loops are associated with a high level of rearrangement mutations, as shown in this study. Other studies indicate that microinjected heteroduplexes with single-base mismatches are associated with a higher than expected mutation frequency (21). Even the presence of a nick, which is assumed to be the first step in repair of mismatches, apparently induces a high frequency of local point mutations (22). These observations, coupled to the very high frequency of mutation induced by heteroduplex formation in gene targeting experiments (23), suggest that cultured mammalian cells lack fidelity in correcting mismatched DNA.

ACKNOWLEDGMENTS

We thank De Dieu, Kathleen Marburger, and Yue-Xian Hou for expert technical assistance. We are grateful to Raju Kucherlapati and Michael Seidman for stimulating discussions and for sharing their data prior to publication. This work was supported by grants from the Public Health Service (CA15743 and GM33405) and the Robert A. Welch Foundation (Q-977).

REFERENCES

1. Holliday, R. (1965) *Genet. Res.* 5, 282-304.
2. Meselson, M. and Radding, C. (1975) *Proc. Natl. Acad. Sci. USA* 72, 358-361.
3. Orr-Weaver, T.L. and Szostak, J.W. (1985) *Microbiol. Rev.* 49, 33-58.
4. Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J. and Stahl, F.W. (1983) *Cell* 33, 25-35.
5. Bianchi, M.E. and Radding, C.M. (1983) *Cell* 35, 511-520.
6. Hamza, H., Kalogeropoulos, A. and Rossignol, J.-L. (1986) *Proc. Natl. Acad. Sci. USA* 83, 7386-7390.
7. Lichten, M. and Fox, M.S. (1983) *Genetics* 103, 5-22.
8. Lichten, M. and Fox, M.S. (1984) *Proc. Natl. Acad. Sci. USA* 81, 7180-7184.
9. Radding, C.M. (1978) *Ann. Rev. Biochem.* 47, 847-880.
10. White, J.H., Lusnak, K. and Fogel, S. (1985) *Nature* 315, 350-352.
11. White, R.L. and Fox, M.S. (1974) *Proc. Natl. Acad. Sci. USA* 71, 1544-1548.
12. Zagursky, R.J., Baumeister, K., Lomax, N. and Berman, M.L. (1985) *Gene Anal. Tech.* 2, 89-94.
13. Weiss, U. and Wilson, J.H. (1987) *Proc. Natl. Acad. Sci. USA* 84, 1619-1623.
14. Hirt, B. (1967) *J. Mol. Biol.* 26, 365-369.
15. Roth, D.B. and Wilson, J.H. (1986) *Mol. Cell. Biol.* 6, 4295-4304.
16. Roth, D.B., Porter, T.N. and Wilson, J.H. (1985) *Mol. Cell. Biol.* 5, 2599-2607.
17. Hauser, J., Seidman, M.M., Sidur, K. and Dixon, K. (1986) *Mol. Cell. Biol.* 6, 277-285.
18. Miller, J.H., Lebkowski, J.S., Greisen, K.S. and Calos, M.P. (1984) *EMBO J.* 3, 3117-3121.

19. Sarkar, S., DasGupta, U.B. and Summers, W.C. (1984) *Mol. Cell. Biol.* 4, 2227-2230.
20. Seidman, M.M., Dixon, K., Razzaque, A., Zagursky, R.J. and Berman, M.L. (1985) *Gene* 38, 233-237.
21. Folger, K.R., Thomas, K. and Capecchi, M.R. (1985) *Mol. Cell. Biol.* 5, 70-74.
22. Seidman, M., Bredberg, Anders, Seetharam, S., and Kraemer, K.H. (1987) *Proc. Natl. Acad. Sci. USA* 84, 4944-4948
23. Thomas, K.R. and Capecchi, M.R. (1986) *Nature* 324, 34-38.
24. Thimmappaya, B. and Shenk, T. (1979) *J. Virol.* 30, 668-673.
25. Van Heuverswyn, H. and Fiers, W. (1979) *Eur. J. Biochem.* 100, 51-60.
26. Buchman, A.R., Burnett, L. and Berg, P. (1980) In Tooze, J. (ed). *DNA Tumor Viruses*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.