Copyright © 1999 by the Genetics Society of America

Suppression of Intrachromosomal Gene Conversion in Mammalian Cells by Small Degrees of Sequence Divergence

Tamas Lukacsovich and Alan S. Waldman

Department of Biological Sciences, University of South Carolina, Columbia, South Carolina 29208

Manuscript received August 7, 1998
Accepted for publication December 14, 1998

ABSTRACT

Pairs of closely linked defective herpes simplex virus (HSV) thymidine kinase (*tk*) gene sequences exhibiting various nucleotide heterologies were introduced into the genome of mouse Ltk$^-$ cells. Recombination events were recovered by selecting for the correction of a 16-bp insertion mutation in one of the *tk* sequences. We had previously shown that when two *tk* sequences shared a region of 232 bp of homology, interruption of the homology by two single nucleotide heterologies placed 19 bp apart reduced recombination nearly 20-fold. We now report that either one of the nucleotide heterologies alone reduces recombination only about 2.5-fold, indicating that the original pair of single nucleotide heterologies acted synergistically to inhibit recombination. We tested a variety of pairs of single nucleotide heterologies and determined that they reduced recombination from 7- to 175-fold. Substrates potentially leading to G-G or C-C mispairs in presumptive heteroduplex DNA (hDNA) intermediates displayed a particularly low rate of recombination. Additional experiments suggested that increased sequence divergence causes a shortening of gene conversion tracts. Collectively, our results suggest that suppression of recombination between diverged sequences is mediated via processing of a mispaired hDNA intermediate.

TWO approaches have been used to study the homology requirements of recombination in mammalian cells. One approach involves measuring the rate of recombination as a function of the length of homology shared by two sequences. In mammalian cells, the rate of intrachromosomal recombination between two closely linked sequences drops off sharply when the amount of homology is reduced from 295 to 200 bp (Liskay *et al.* 1987). Efficient extrachromosomal recombination in mammalian cells was also shown to require ~200 bp of homology (Rubnitz and Subramani 1984). A second approach is to measure the effect of nucleotide mismatches on the rate of recombination. We have shown that intrachromosomal recombination in mammalian cells between two linked sequences sharing 81% homology was reduced over 1000-fold relative to recombination between sequences displaying near-perfect homology (Waldman and Liskay 1987). Further studies suggested that efficient intrachromosomal recombination in mammals requires a minimum target of between 134 and 232 bp of continuous homology (Waldman and Liskay 1988). In studies of recombination in bacteria, Shen and Huang (1986, 1989) coined the phrase "minimal efficient processing segment" (MEPS) to describe the minimal length of perfect homology needed for efficient recombination. A model was developed for bacteria (Shen and Huang 1986) in which the recombination rate is essentially proportional to the total number of MEPS units shared by two sequences. The MEPS value for the RecBCD-mediated pathway of *Escherichia coli* acting in phage-plasmid recombination was determined to be ~30 bp (Shen and Huang 1986, 1989). Using similar terminology, we estimated the MEPS value for intrachromosomal homologous recombination in mammalian cells to be between 134 and 232 bp (Waldman and Liskay 1988).

Intrinsic to the original concept of MEPS as developed for recombination in bacteria is the notion that the minimal target required for recombination initiation consists of a segment of perfect, uninterrupted homology and can be perturbed by a single nucleotide heterology. In our earlier estimation of the value of MEPS in mouse fibroblasts (Waldman and Liskay 1988), intrachromosomal recombination was reduced 20-fold when a 232-bp homology interval was interrupted by two single nucleotide heterologies separated by 19 bp. It thus became clear that intrachromosomal recombination in mammalian cells can be exquisitely sensitive to very small degrees of sequence divergence. However, the question of whether a single nucleotide heterology represents sufficient sequence divergence to disrupt recombination was not formally tested.

In the current work, we refined our evaluation of the minimal homology requirements for intrachromosomal recombination in mammalian cells. We examined whether or not mammalian recombination machinery is sensitive to a single nucleotide heterology and whether

Corresponding author: Alan S. Waldman, Department of Biological Sciences, University of South Carolina, 700 Sumter St., Columbia, SC 29208. E-mail: awaldman@sc.edu

or not different specific nucleotide heterologies exert differential effects on intrachromosomal recombination. In this report we show that a single nucleotide heterology placed within MEPS is sufficient to reduce recombination by about 2.5-fold, whereas a pair of single nucleotide heterologies can act together to significantly reduce recombination from 7- to 175-fold. Different combinations of nucleotide heterologies inhibit recombination to different extents, with heterologies leading to the formation of G-G or C-C mispairs in presumptive heteroduplex DNA (hDNA) intermediates being particularly inhibitory. Our data suggest that recombination between diverged sequences in mammalian cells is suppressed at the level of cellular processing of a mismatched hDNA intermediate. Other consequences of sequence divergence on recombination are also discussed.

MATERIALS AND METHODS

Cell culture and derivation of experimental cell lines: Mouse L cells deficient in thymidine kinase (L*tk*⁻ cells) were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum, 0.1 mm minimal essential amino acids (GIBCO, Gaithersburg, MD), and 50 μg/ml gentamicin sulfate. To select for tk^+ cells, growth medium was supplemented with hypoxanthine/aminopterin/thymidine (HAT; GIBCO). Cells were maintained at 37° in a humidified atmosphere of 5% CO_2.

Plasmid DNA was linearized by cleavage with *Cla*I and introduced into mouse Ltk^- cells by nuclear microinjection (Capecchi 1980) or electroporation (Lukacsovich *et al.* 1994). Stable transformants were isolated following selection in G418 (200 μg/ml active drug) as previously described (Waldman and Liskay 1987).

Plasmid descriptions: All plasmids are based on the vector pJS-1, which is a derivative of pSV2neo (Southern and Berg 1982) with restriction site modifications as previously described (Liskay *et al.* 1984). The #28 mutant herpes simplex virus type 1 (HSV-1) thymidine kinase (*tk*) gene contains an 8-bp *Xho*I linker inserted at nucleotide 1036 [numbering according to Wagner *et al.* (1981)] of the *tk* gene from HSV-1 strain F. Plasmid pTK28 (Waldman and Liskay 1987) contains the #28 mutant *tk* gene on a 2.5-kb *Bam*HI fragment inserted into the unique *Bam*HI site of the vector. A 360-bp fragment (nucleotide positions 808–1168) of the *tk* gene from HSV-1 strain 101 was inserted into the unique *Hin*dIII site of vector sequences on pTK28 to produce pAG-CT, previously referred to as pF-101 (Waldman and Liskay 1988). Plasmid pAA-CC, previously referred to as pF-F (Waldman and Liskay 1988), is identical to pAG-CT except that the 360-bp fragment inserted into the *Hin*dIII site originated from the *tk* gene from HSV-1 strain F. To simplify nomenclature, the names assigned to the plasmids described in this work contain the specific nucleotide differences displayed by the pair of *tk* sequences contained on the plasmid. For example, "pAG-CT" contains *tk* sequences displaying an A/G heterology at position 942 and a C/T heterology at position 962 of the *tk* coding region (in addition to the *Xho*I linker insertion heterology at position 1036). The *tk* genes on pAA-CC display no heterologies other than the *Xho*I linker insertion. Nucleotide heterologies described refer to sequence differences on the *tk* coding strand, which we define as the DNA strand equivalent to the mRNA sequence, and are hereafter to be understood to exist in addition to the *Xho*I linker insertion mutation. The nucleotide heterologies contained on pAG-CT and other plasmids described below are indicated in the schematic presented in Figure 1c.

Site-directed mutagenesis, using the Transformer site-directed mutagenesis kit (Clonetech, Palo Alto, CA), was used in conjunction with other standard molecular biological techniques to construct six additional plasmid substrates (Figure 1c). Plasmid pAA-TC contains a pair of *tk* sequences displaying a T/C heterology at position 962, while pAG-TT contains *tk* sequences with an A/G heterology at position 942 of the *tk* coding region. In addition to an A/G heterology at position 942, the pairs of *tk* sequences on pAG-TG, pAG-CG, pAG-AG, and pAG-AT additionally display, respectively, a T/G, C/G, A/G, and an A/T mismatch at position 959.

Determination of intrachromosomal recombination frequencies: Starting with 10 single cells, 10 independent subclones were generated from each cell line. The subclones were propagated to 25–40 × 10^6 cells each and then plated separately into HAT medium to select for tk^+ segregants. Recombination frequency for each cell line was calculated by dividing the total number of HAT^R colonies (for all 10 subclones combined) by the total number of copies of integrated DNA substrate plated. The mean recombination frequency for *all* cell lines containing a particular substrate was calculated by dividing the total number of HAT^R colonies by the total number of copies of integrated DNA substrate plated for all pertinent cell lines combined. The 95% confidence limits for mean recombination frequency were estimated by making the approximation that numbers of HAT^R colonies recovered from a series of independent subcultures can be described by a Poisson distribution. This approximation is particularly apt for populations yielding small numbers of colonies (Luria and Delbrück 1943). The total number of colonies obtained for cell lines containing a particular DNA substrate can be taken as the mean, which equals the variance, of a Poisson distribution. The square root of variance gives standard deviation, and doubling the standard deviation gives an approximation to the 95% confidence limits for colony number. Dividing this value by the total number of copies of integrated DNA substrate plated gives an estimate for the 95% confidence limits for recombination frequency, as reported in Table 1.

Genomic DNA isolation and Southern blotting: Genomic DNA was prepared from cultured cells and analyzed by Southern hybridization using a ^{32}P-labeled probe specific for the HSV-1 *tk* sequence as previously described (Letsou and Liskay 1987).

PCR amplification and DNA sequence determination: HSV-1 *tk* sequences were PCR amplified from genomic DNA isolated from HAT^R colonies using the following set of primers: 5′-TCTACACCACACAACACCGC-3′ and 5′-ACAAACGACCCAA CACCCGT-3′. The first primer is the sequence of the coding strand of the HSV-1 *tk* gene (strain 101 or F) from nucleotide 814 to 833, while the second primer is the noncoding sequence from nucleotide position 1734 to 1715. PCR was carried out in a 100-μl reaction cocktail containing 333 ng genomic DNA, 1.5 μm of each primer, 10 mm Tris-HCL, pH 8.3, 0.5% gelatin, 50 nm KCl, 1.1 mm $MgCl_2$, 200 μm of each of the four dNTPs, 2.5 units of AmpliTaq™ DNA polymerase (Perkin Elmer Cetus, Norwalk, CT). Each reaction cycle consisted of 1 min at 95°, 1 min at 62°, 3 min at 72°. PCR amplification was accomplished by 35 reaction cycles.

PCR products (920 bp in length) were purified by phenol extraction and ethanol precipitation, cleaved with *Sph*I and *Msc*I, cloned into M13mp18, and sequenced from single-stranded DNA templates using a Sequenase Version 2.0 DNA sequencing kit (Amersham Life Science, Inc., Cleveland).

RESULTS

Recombination is modestly reduced by the introduction of a single nucleotide heterology into the minimal recombination target in mammalian cells: Our previous studies indicated that the value for MEPS for intrachromosomal recombination in mouse cells is between 134 and 232 bp and can be disrupted by two single nucleotide heterologies separated by 19 bp (Waldman and Liskay 1988). We observed that two nucleotide mismatches within MEPS reduced recombination 20-fold. To explore the question of whether recombination in a mammalian cell can be disrupted by a *single* nucleotide heterology, the recombination substrates pAA-TC and pAG-TT (Figure 1c) were produced and stably introduced into the genome of mouse L*tk*$^-$ cells. Two cell lines containing a single integrated copy of each construct were isolated, and the frequencies of intrachromosomal recombination between the pairs of *tk* sequences on the integrated constructs were determined (Table 1). For all the experiments discussed in this work, only gene conversions, but not crossovers, were recoverable. A crossover would produce a nonfunctional *tk* gene with a 5′ or 3′ truncation because the 360-bp *tk* fragment on each construct (Figure 1a) had 5′ and 3′ deletions of coding sequence. It is also to be understood that, in addition to the specified nucleotide mismatch heterologies, the pairs of *tk* sequences on all constructs displayed an 8-bp insertion heterology due to the *Xho*I linker insertion mutation in one *tk* sequence (Figure 1a). We selected for correction of the linker insertion mutation in our experiments.

As indicated in Table 1, the frequencies of *tk*-positive segregants for cell lines containing pAA-TC (mean frequency equal to 7.5×10^{-8}) or pAG-TT (mean frequency equal to 6.7×10^{-8}) were comparable to one another. When corrected for the nonrecombinants among the recovered colonies (see below) these frequencies (5.5×10^{-8} for pAA-TC and 5.7×10^{-8} for pAG-TT) are 2.5-fold lower than the frequency of 1.4×10^{-7} for cell lines containing pAA-CC (Waldman and Liskay 1988; Table 1), which contains *tk* sequences displaying no mismatches (Figure 1c). The recombination frequencies for lines containing pAA-TC and pAG-TT were greater than the frequency of 9×10^{-9} for cell lines containing pAG-CT (Waldman and Liskay 1988; Table 1), which contains a pair of *tk* sequences with two nucleotide heterologies (Figure 1c). The results obtained for lines containing pAG-TT were particularly interesting because the placement of the single A/G heterology in pAG-TT retained the longest stretch of homology of 134 bp, which was also present in the pAG-CT substrate containing two nucleotide heterologies (Figure 1b), and yet the recombination frequency was increased by about 7-fold in pAG-TT cell lines relative to the pAG-CT cell lines. These results suggest that although the introduction of a single nucleotide heterology into MEPS can reduce intrachromosomal homologous recombination by about 2.5-fold, a *pair* of nucleotide heterologies is more effective at reducing recombination.

Southern blotting analysis of HAT^R segregants arising from recombination between two *tk* sequences displaying a single nucleotide heterology: DNA samples were isolated from 19 HAT^R segregants from cell lines containing pAA-TC and 42 HAT^R segregants arising from cell lines containing pAG-TT. Southern blotting analysis was performed using a probe specific for herpes simplex virus (HSV) *tk* sequences; a representative analysis is presented in Figure 2. In all, 36 out of 42 HAT^R clones examined from lines containing pAG-TT displayed a 2.5-kb *Bam*HI fragment that was resistant to cleavage by *Xho*I (Figure 2, lanes 1, 2, 3, 5, 6), as expected for a gene conversion event that corrected the *Xho*I linker insertion mutation (see Figure 1a). The remaining six HAT^R clones unexpectedly displayed a 2.5-kb *Bam*HI fragment that was cleavable by *Xho*I into 1.3-kb and 1.2-kb fragments (Figure 2, lane 4). For HAT^R clones derived from cell lines containing pAA-TC, 14 out of 19 displayed a 2.5-kb *Bam*HI fragment resistant to *Xho*I (Figure 2, lanes 7–11), while the remaining 5 clones displayed a 2.5-kb *Bam*HI fragment cleavable by *Xho*I (Figure 2, lane 12). HAT^R clones that retained the *Xho*I linker insertion mutation presumably arose from the introduction of a second, compensating mutation in the *tk* gene. Results from Southern blotting analyses are summarized in Table 2. These data were used to calculate the corrected recombination frequencies of 5.5×10^{-8} and 5.7×10^{-8} for pAA-TC and pAG-TT cell lines (Table 1) by multiplying the measured frequency of HAT^R segregants for each substrate by the fraction of segregants that had actually undergone a gene conversion.

A pair of single nucleotide heterologies can further disrupt efficient recombination in mammalian cells: The above experiments indicated that a single nucleotide heterology introduced into a MEPS interval of homology reduced intrachromosomal recombination by about 2.5-fold, whereas previous studies (Waldman and Liskay 1988) indicated that a pair of nucleotide heterologies reduced recombination by nearly 20-fold. We wanted to determine if recombination is equally sensitive to various combinations of nucleotide heterologies. To this end, we constructed cell lines from mouse L*tk*$^-$ cells containing plasmids pAG-TG, pAG-CG, pAG-AG, or pAG-AT (Figure 1c) stably integrated into the genome. These DNA constructs each contain a pair of *tk* sequences displaying an A/G heterology plus one additional single nucleotide heterology. Frequencies of *tk*$^+$ segregants were determined for at least two cell lines for each recombination substrate (Table 1). As indicated in Table 1, frequencies for cell lines containing pAG-TG, pAG-CG, pAG-AG, or pAG-AT were reduced relative

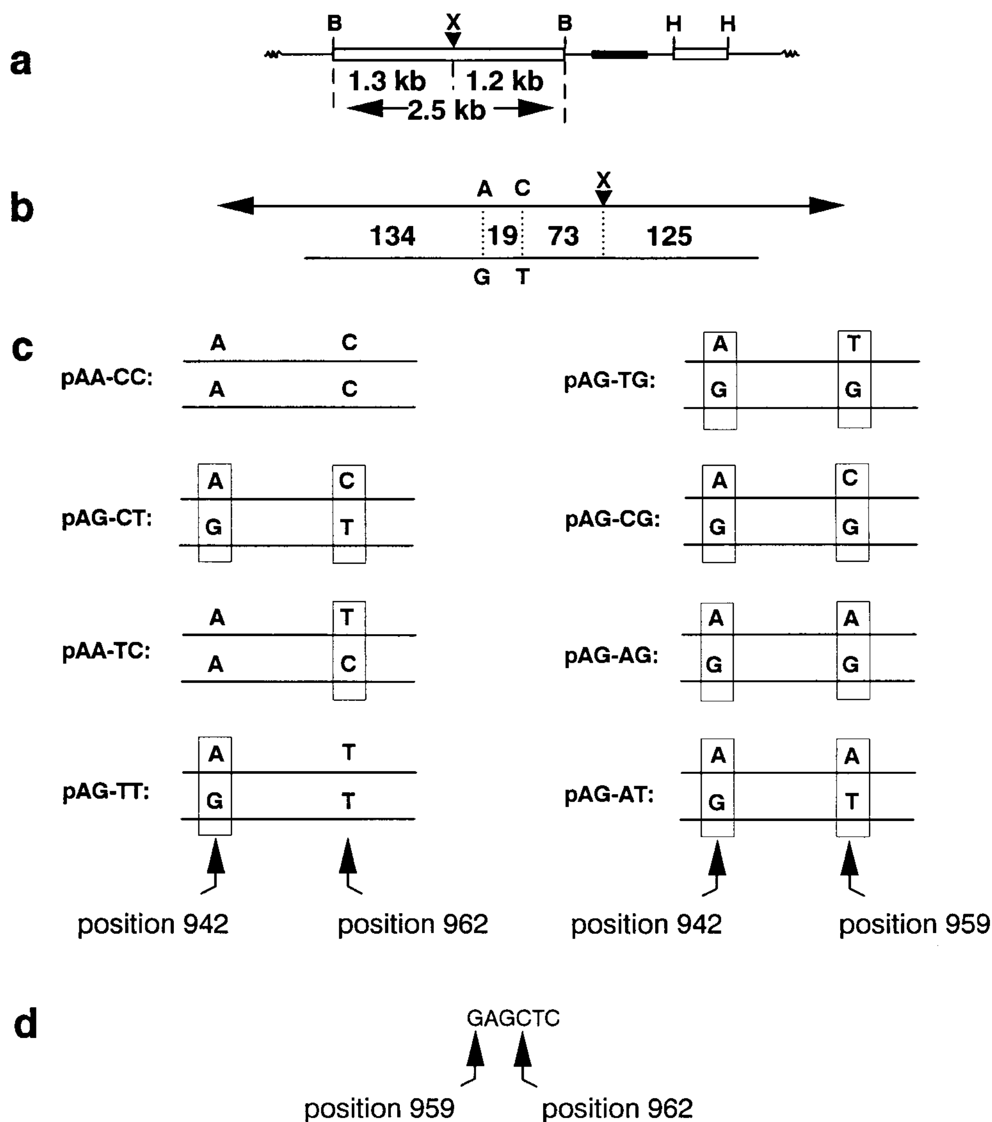

Figure 1.—Recombination substrates. (a) Shown is a schematic of a generic DNA construct integrated into the genome of a mouse L*tk*⁻ cell. Constructs are linearized with *Cla*I prior to introduction into cells. The *Cla*I site is likely destroyed upon integration and is not shown. Stable integrants are selected based on G418^R conferred by the *neo* gene (black rectangle). Cloned between *Bam*HI (B) sites is a 2.5-kb fragment containing a *Xho*I (X) linker insertion mutant *tk* gene, while cloned between *Hin*dIII (H) sites is a 360-bp internal fragment of the *tk* gene. (b) Shown is a schematic of the relevant portion of coding strands of the *tk* sequences from construct pAG-CT aligned in homologous register, with the coding strand from the 360-bp *tk* fragment on the bottom. Indicated are the positions of the A/G and C/T heterologies and the *Xho*I linker insertion as well as the lengths, in base pairs, of stretches of uninterrupted homologies. (c) Depicted are the pairs of *tk* sequences in the various constructs that display up to two single nucleotide heterologies (in addition to the *Xho*I linker insertion mutation). The heterologies, shown in boxes, represent differences in coding strand sequences, with the upper line in each schematic representing the sequence of the insertion mutant *tk* gene and the bottom line representing the sequence of the 360-bp *tk* fragment. The positions of heterologies in all constructs are similar to those indicated for pAG-CT in b, except that one heterology in constructs pAG-TG, pAG-CG, pAG-AG, and pAG-AT is located at position 959 rather than 962, as indicated at the bottom of the schematic. (d) Nucleotide sequence of the coding strand of the wild-type (strain F) *tk* gene from position 959–964. The sequence shown, 5′-GAGCTC-3′, is a recognition site for *Sac*I, which is disrupted by base substitutions at position 959 or 962 in several *tk* sequences in the DNA constructs.

to recombination frequencies for cell lines containing pAA-TC or pAG-TT. Three cell lines containing pAG-CG produced no HAT^R segregants (for an average recombination frequency of less than 8×10^{-10}), while the cell lines harboring each of the other constructs containing two nucleotide heterologies displayed average recombination frequencies ranging from 1.2×10^{-8} to 2.0×10^{-8}, once corrected for nonrecombinant tk^+ revertants (see below). These latter frequencies were roughly 3- to 5-fold lower than the average recombination frequency of 5.6×10^{-8} for lines containing pAA-TC or pAG-TT (Table 1) and similar to the recombination frequency of 9×10^{-9} previously determined for cell lines containing pAG-CT (Waldman and Liskay 1988; Table 1). These results indicated that a pair of single nucleotide heterologies exerted an additive or modestly synergistic negative effect on recombination because, relative to a substrate with no heterologies, recombination frequency was reduced about 2.5-fold by a single nucleotide heterology and was reduced an additional 3- to 5-fold (or more) by the introduction of a second nucleotide heterology. Two nucleotide heterologies had a combined effect of reducing recombination about 7- to 175-fold (Table 1).

Analysis of HAT^R segregants arising from recombination between a pair of *tk* sequences displaying a pair of

TABLE 1

Frequencies of intrachromosomal recombination between mismatched sequences

DNA substrate[a]	Cell line	Copy no.[b]	No. of cells tested ($\times 10^{-6}$)[c]	No. of HAT^R colonies	Recombination frequency ($\times 10^{8}$)[d]
pAA-CC	1	1	226	52	23
	2	2	250 (500)	53	10
	3	2	270 (540)	73	13
					Mean = 14 ± 2.1
pAG-CT	1	1	280	0	<0.4
	2	1	280	1	0.4
	3	2	260 (520)	9	1.7
					Mean = 0.9 ± 0.6
pAA-TC	1	1	342	10	2.9
	2	1	349	42	10.0
					Mean = 7.5 ± 2.1
					Corrected = 5.5
pAG-TT	1	1	337	25	7.4
	1	1	522	34	6.5
	2	1	441	28	6.3
					Mean = 6.7 ± 1.4
					Corrected = 5.7
pAG-TG	1	1	358	10	2.8
	2	1	356	4	1.1
					Mean = 2.0 ± 1.0
pAG-CG	1	1	356	0	—
	2	1	430	0	—
	3	1	430	0	—
					Mean < 0.08
pAG-AG	1	1	288	4	1.4
	2	1	349	8	2.3
					Mean = 1.9 ± 1.1
					Corrected = 1.6
pAG-AT	1	1	337	5	1.5
	2	1	354	5	1.4
					Mean = 1.5 ± 0.9
					Corrected = 1.2

[a] Substrates are as described in materials and methods and depicted schematically in Figure 1. Raw data for cell lines containing pAA-CC and pAG-CT were previously reported (Waldman and Liskay 1988).
[b] Number of copies of relevant DNA substrate in the particular cell line.
[c] The total number of copies of integrated substrate plated is given in parentheses.
[d] Number of HAT^R colonies divided by total number of copies of integrated DNA substrate plated for each cell line. The *mean* recombination frequencey is the total number of HAT^R colonies divided by total number of copies of integrated substrate plated for all cell lines containing a particular DNA substrate combined; 95% confidence limits are shown. The *corrected* reconbination frequency is the mean recombination frequency multiplied by the fraction of analyzed HAT^R colonies that had actually undergone a gene conversion (see Table 2). Corrected frequencies are presented only for substrates from which one or more nonrecombinant HAT^R segregants were recovered.

single nucleotide heterologies: DNA samples isolated from 20 independent HAT^R segregants from various cell lines containing either pAG-TG, pAG-AG, or pAG-AT were analyzed by Southern blotting using an HSV-1 *tk*-specific probe (Figure 3). In total, 14 out of the 20 clones analyzed displayed a 2.5-kb *Bam*HI fragment resistant to *Xho*I cleavage (Figure 3, lanes 1, 3, 5–8), the pattern expected for correction of the insertion mutation via simple gene conversion (see Figure 1a). One HAT^R segregant from a cell line containing pAG-TG and three HAT^R segregants from a cell line containing pAG-AT displayed both a 2.5-kb *Bam*HI fragment that resisted *Xho*I cleavage as well as a 2.5-kb *Bam*HI fragment that cleaved with *Xho*I (for example, Figure 3, lanes 2 and 10). These exceptional recombinants had undergone an increase in *tk* gene copy number and may have arisen from nondisjunction (see discussion). One HAT^R clone arising from a cell line containing pAG-AG and one clone arising from a cell line containing pAG-AT displayed a single 2.5-kb *Bam*HI fragment that retained the *Xho*I site (Figure 3, lanes 4 and 9). The sequence in the vicinity of the *Xho*I linker was determined for

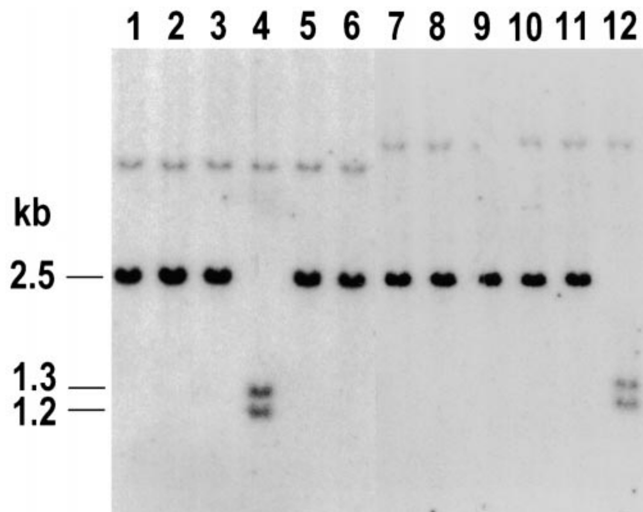

Figure 2.—Southern blotting analysis of HAT[R] segregants arising from recombination between *tk* sequences displaying a single nucleotide heterology. Genomic DNA isolated from HAT[R] segregants from cell lines containing pAG-TT (lanes 1–6) or pAA-TC (lanes 7–12) were subjected to Southern blotting analysis using an HSV-1 *tk*-specific probe. Shown are 6 representative samples out of a total of 61 segregants analyzed. DNA (8 μg) from each HAT[R] segregant was digested with *Bam*HI alone (odd lanes) or *Bam*HI plus *Xho*I (even lanes) and displayed in a pair of adjacent lanes. The origins of the 2.5-, 1.3-, and 1.2-kb bands are illustrated at the top of Figure 1. The additional band in all lanes is a junction fragment that is detected due to the homology between the probe used and the 360-bp *tk* fragment on pAG-TT and pAA-TC. See text for further discussion.

the exceptional HAT[R] clone isolated from the cell line containing pAG-AG. This clone had an 8-bp deletion mapping immediately downstream from the *Xho*I recognition site in the linker insertion, restoring the reading frame of the *tk* gene (data not shown). A summary of the data obtained from Southern blotting analysis is presented in Table 2.

The observation of a mutated *tk* gene in a HAT[R] segregant raised the possibility that the *tk* sequence was not wild type even in the majority of the HAT[R] recombinants that displayed a 2.5-kb *Bam*HI fragment that was resistant to *Xho*I cleavage. To address this issue, we took advantage of the fact that the *Xho*I linker insertion mutation disrupted an *Eag*I restriction site in the *tk* gene. Reconstruction of a wild-type *tk* sequence would therefore have restored a functional *Eag*I site. A 920-bp DNA sequence encompassing the original position of the *Xho*I linker insertion mutation was PCR amplified from genomic DNA from six independent HAT[R] recombinants from pAG-TG cell lines and from four independent HAT[R] recombinants from pAG-AG cell lines, and the PCR products were digested with *Eag*I. All 10 PCR products cleaved with *Eag*I at the site where the *Xho*I linker had been positioned (data not shown). These results demonstrated that the wild-type *tk* sequence had been restored. In addition, the nucleotide sequences were determined for *tk* genes isolated from two recombinants from pAG-TG lines and two recombinants from pAG-AG lines, and in each case the sequence was wild type (data not shown).

Shortening of conversion tracts with increasing degree of sequence divergence: We made an assessment of the length of conversion tracts for recombinants isolated from cell lines containing pAG-TG or pAG-AG. As mentioned above, we determined the nucleotide sequences for four such recombinants. This analysis revealed that neither the single nucleotide heterology at position 942 nor the heterology at position 959 had been coconverted along with the corrected *Xho*I linker insertion. Additionally, we PCR amplified a 920-bp sequence of the functional *tk* sequence surrounding the original position of the *Xho*I linker insertion from six independent recombinants from pAG-TG lines and from four independent recombinants from pAG-AG lines and digested these 920-bp PCR products with *Sac*I.

TABLE 2

Summary of results from Southern blotting analyses of recombinants

DNA substrate	No. of recombinants analyzed	No. of gene conversions[a]	No. of mutants[b]	No. of duplications[c]	Fraction coconverted[d]
pAA-TC	19	14	5	0	5/10
pAG-TT	42	36	6	0	ND
pAG-TG	8	7	0	1	0/8
pAG-AG	7	6	1	0	0/6
pAG-AT	5	1	1	3	ND

[a] Number of HAT[R] segregants that had the expected pattern for a simple gene conversion.

[b] Number of HAT[R] segregants that did not lose the *Xho*I linker insertion and presumably acquired a second compensating mutation.

[c] Number of HAT[R] segregants that, in addition to having undergone a gene conversion, displayed an increase in copy number of the recombination substrate possibly resulting from nondisjunction or amplification.

[d] First number (before the slash) is the number of clones that showed coconversion of the *Sac*I site along with correction of the *Xho*I linker insertion mutation. The second number is the number of clones tested. ND, not done.

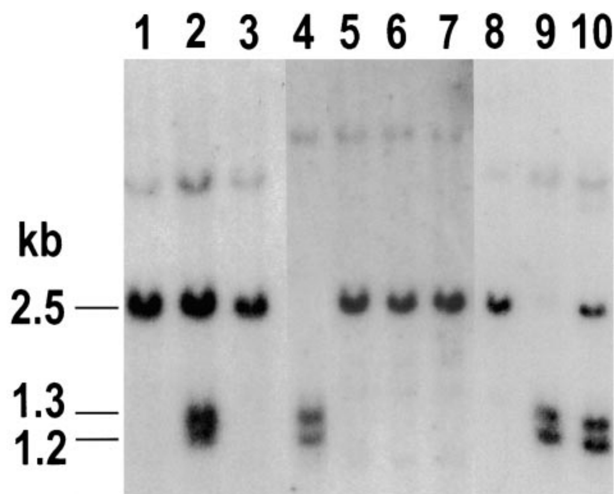

Figure 3.—Southern blotting analysis of HAT^R segregants arising from recombination between *tk* sequences displaying two heterologous nucleotides. Genomic DNA isolated from HAT^R segregants from cell lines containing pAG-TG (lanes 1–3), pAG-AG (lanes 4–7), and pAG-AT (lanes 8–10) were digested with *Bam*HI plus *Xho*I and subjected to Southern blotting analysis using an HSV-1 *tk*-specific probe. Shown are 10 representative samples out of 20 segregants analyzed. The origins of the 2.5-, 1.3-, and 1.2-kb bands are illustrated at the top of Figure 1. Additional higher molecular weight bands are junction fragments. See text for further discussion.

Had the heterology at position 959 (the mismatch closest to and 76 bp upstream from the *Xho*I linker insertion at position 1036) been coconverted along with the corrected *Xho*I linker insertion, then a *Sac*I site would have been restored to the recipient *tk* sequence because the 360-bp donor sequence contained a *Sac*I site at this position (see Figure 1d). None of the 10 PCR fragments were cleavable with *Sac*I (Table 2), indicating that the closer of the nucleotide heterologies had not been coconverted with the selected marker in any of the recombinants analyzed. We also analyzed recipient *tk* genes from 10 independent recombinants recovered from cell lines containing pAA-TC, and containing a single nucleotide heterology at position 962 (Figure 1c and d), and determined that 5 of the 10 recombinants displayed a coconversion of the heterologous nucleotide along with the corrected *Xho*I linker insertion mutation (Table 2). These latter results were consistent with previously published work (Liskay and Stachelek 1986) indicating that a silent genetic marker was coconverted with a selected marker 85 bp away in ~50% of intrachromosomal gene conversions in mouse L*tk*⁻ cells. The difference we observed between the coconversion frequencies for substrates containing two heterologous nucleotides *vs.* substrates containing a single heterologous nucleotide is statistically significant ($P < 0.0011$ by a *g*-test; Sokal and Rohlf 1981). Our coconversion analysis suggests a shortening of gene conversion tracts as the number of heterologous nucleotides increased from one (in pAA-TC) to two (in pAG-TG or pAG-AG).

DISCUSSION

Homologous recombination in mammalian cells is sensitive to both the length and degree of homology shared by two sequences (Rubnitz and Subramani 1984; Liskay *et al.* 1987; Waldman and Liskay 1987, 1988; Yang and Waldman 1997). We had previously determined (Waldman and Liskay 1988) that the MEPS value for intrachromosomal recombination in mouse fibroblasts is between 134 and 232 bp and that a pair of single nucleotide heterologies embedded within a 232-bp stretch of homology was sufficient to disrupt MEPS and bring about nearly a 20-fold reduction in recombination rate. In those earlier studies, we conjectured that a single nucleotide heterology would likely be sufficient to disrupt MEPS, consistent with studies of bacterial recombination (Shen and Huang 1986, 1989) indicating a requirement for *uninterrupted* homology. In the current work we observed that a single nucleotide heterology is indeed sufficient to reduce recombination, but only by about 2.5-fold, while a pair of nucleotide heterologies can act to suppress recombination from 7-fold to as much as 175-fold. We have subsequently been forced to partially rethink our concept of MEPS as it applies to intrachromosomal recombination in mammalian cells. We previously speculated (Waldman and Liskay 1988) that scattered single nucleotide heterologies would affect the recombination rate significantly if they were distributed so that the greatest distance between consecutive heterologies were below the critical value of MEPS. Our current work demonstrates that recombination rate is not simply a function of the length of the longest stretch of perfect homology along a given DNA interval since recombination was seen to be more sensitive to a pair of single nucleotide heterologies than to a single nucleotide heterology, irrespective of the length of the longest stretch of homology in a defined DNA interval.

All tested pairs of nucleotide heterologies reduced recombination significantly, but not all mismatched bases exerted equal effects on recombination levels (Table 1). Implicit in the very observation that different combinations of nucleotide heterologies have different effects on recombination is the inference that heterologous nucleotides impact recombination at a step *after* the formation of a mismatched hDNA intermediate and that differential inhibitory effects likely are related to differential processing, or repair, of specific base-base mispairs. Related to this point is the report (de Wind *et al.* 1995) that inactivation of the *Msh2* DNA mismatch repair gene in mice alleviates the sensitivity of targeted homologous recombination to very small degrees of sequence divergence. Other studies in yeast (Negritto *et al.* 1997) also indicated that mismatches can interfere with targeted gene replacement only if they are included in hDNA formed between the recombining molecules. Our cell lines containing the substrate pAG-CG dis-

played an undetectably low frequency of recombination (see Table 1). During recombination leading to gene conversion, hDNA can theoretically be formed between either the coding strand of the recipient *tk* gene and the noncoding strand of the donor *tk* sequence, or vice versa. Depending on which type(s) of heteroduplex forms, either a G-G or a C-C mispair (or both) would be generated during recombination in cell lines containing pAG-CG. It has been demonstrated that all eight possible base-base mispairs are repaired in monkey kidney cells (Brown and Jiricny 1988) and Chinese hamster ovary cells (Bill *et al.* 1998). A preferential retention of the bases G and C over A and T during mismatch repair was observed (Bill *et al.* 1998; Brown and Jiricny 1988). It is intriguing to speculate that G-G or C-C mispairs (generated in hDNA in pAG-CG cell lines) are subject to particularly efficient mismatch recognition and/or processing that has the effect of efficiently blocking the successful execution of recombination.

In addition to exhibiting a reduced recombination rate, substrates with *tk* sequences possessing two heterologous nucleotides also appeared to generate gene conversion tracts that were shorter than tracts generated with substrates possessing only a single heterologous nucleotide. This finding is consistent with work done in yeast (Harris *et al.* 1993) indicating a shortening of conversion tracts as sequence divergence increases. Such an observation is consistent with the notion of "heteroduplex rejection" (Alani *et al.* 1994). In such a paradigm, hDNA formation between sequences displaying low levels of divergence is disallowed. Theoretically, heteroduplex rejection may be accomplished by an inability of branch migration to proceed through mismatched bases (Worth *et al.* 1994), by destruction of "highly" mispaired heteroduplex DNA recombination intermediates (possibly by overlapping mismatch repair tracts; Rayssiguier *et al.* 1989), or by a reversal of branch migration so as to undo highly mispaired heteroduplex (Alani *et al.* 1994; Štambuk and Radman 1998). We cannot definitively distinguish between these three possibilities, which are not mutually exclusive. However, for several reasons we consider it unlikely that branch migration during propagation of a recombination event in mouse cells is blocked by one or two nucleotide heterologies. First, as discussed above, the differential effect of different nucleotide heterologies would appear to be most readily explained by differential processing of specific base-base mispairs, which necessitates that branch migration proceed through heterologies to establish mismatched hDNA. Second, because a *single* nucleotide heterology often does not block branch migration and is frequently included in conversion tracts (see coconversion data for pAA-TC cell lines, Table 2), it is not apparent how the addition of a second heterologous nucleotide would be able to block branch migration through either of the two mismatches. Third, in our previous work (Waldman and Liskay 1988; Yang and Waldman 1997) we recovered products of recombination events in which branch migration apparently proceeded through hundreds of base pairs of sequences displaying only 80% sequence identity, when initiation was allowed to occur within an adjacent interval of 360 bp of perfect homology.

We consider a plausible explanation for our current work to be that the various pairs of heterologous nucleotides tested acted at the very initial steps of recombination during which formation of some critical amount of hDNA (via branch migration) may be required to stabilize joint molecules following strand invasion (Kitts and Nash 1987; Štambuk and Radman 1998). We infer that, relative to a lone mispair, the presence of a *pair* of base-base mispairs in hDNA generated in our substrates was more effective at triggering efficient heteroduplex rejection by destruction or reversal of the hDNA, destabilizing the joint molecule and impeding recombination. Another possibility is that, as found in yeast (Borts and Haber 1987; Borts *et al.* 1990), the presence of multiple base-base mispairs in hDNA may provoke the formation of recombinagenic lesions (by mismatch repair) that may stimulate secondary events and reduce the recovery of viable recombinants. If some or all of these inferences are accurate, then the recombination products that we did recover in the presence of two heterologies may have been generated by a pathway for recombination initiation that circumvented the formation of a significant length of hDNA.

We do not know what influence, if any, the presence of the *Xho*I insertion mutation had on the effect of the point heterologies, but it is not unreasonable to assume that the insertion mutation is recognized differently by cellular repair machinery (Dohet *et al.* 1987; Parker and Marinus 1992; Harris *et al.* 1993; Modrich and Lahue 1996; Tran *et al.* 1996; Sia *et al.* 1997; Taghian *et al.* 1998) and thus its influence on recombination may be independent of the other heterologies. We presently do not know what effect varying the distance between a pair of single nucleotide heterologies would have on recombination nor what effect two consecutive heterologous nucleotides would have on recombination.

We noticed that a surprising percentage of HAT^R segregants retained the original *Xho*I linker insertion mutation (Table 2) and therefore presumably acquired a second compensating mutation. One such mutation was determined by nucleotide sequencing to be a deletion immediately adjacent to the *Xho*I site. Although we cannot conclusively determine whether such mutagenic events were associated with recombinational interactions between *tk* sequences, in earlier work involving cell lines harboring insertion mutant *tk* genes in the absence of a linked homologous *tk* donor sequence we failed to recover any HAT^R segregants from $>10^9$ cells tested (Waldman and Liskay 1987; D. Yang and A. S. Waldman, unpublished results). Additionally, in this

work we did not recover any HAT[R] segregants from three different cell lines containing pAG-CG (Table 1) that also failed to yield any wild-type recombinants. We thus speculate that the mutated HAT[R] segregants recovered in this study are byproducts of recombination attempts between *tk* sequences. We were also surprised to find that a high percentage of recombinants analyzed from cell lines with substrates harboring two nucleotide heterologies seemed to have undergone an apparent duplication of the recombination substrate (see Table 2). Such a change in copy number can result from nondisjunction. It is possible that if recombination occurring between two divergent sequences on sister chromatids generates an intermediate that is slow to resolve, the chromatids may get "stuck together" and fail to segregate properly. It is also possible that the increase in copy number is due to amplification of chromosomal sequences encompassing the integrated construct. We cannot rule out the possibility that the apparent increase in copy number is unrelated to recombination, but we have not observed such copy number change in unselected clones from our cell lines (data not shown). If the increased number of copies of *tk* sequences in some segregants was indeed a byproduct of cellular processing of mismatched hDNA recombination intermediates, the ability of a mammalian cell to suppress recombination between divergent sequences on occasion may come at the cost of provoking additional genetic alterations. It is clear that further elucidation of the molecular players and mechanisms for heteroduplex rejection is needed to develop a more complete understanding of how genomic integrity is maintained in mammalian cells.

We thank Debra Bauerle for her technical expertise and Barbara Criscuolo Waldman for her invaluable advice on the manuscript. This work was supported by U.S. Public Health Service grant GM47110 from the National Institute of General Medical Sciences to A.S.W.

LITERATURE CITED

Alani, E., R. A. G. Reenan and R. D. Kolodner, 1994 Interaction between mismatch repair and genetic recombination in *Saccharomyces cerevisiae.* Genetics **137:** 19–39.

Bill, C. A., W. A. Duran, N. R. Miselis and J. A. Nickoloff, 1998 Efficient repair of all types of single-base mismatches in recombination intermediates in Chinese hamster ovary cells: competition between long-patch and G-T glycosylase-mediated repair of G-T mismatches. Genetics **149:** 1935–1943.

Borts, R. H., and J. E. Haber, 1987 Meiotic recombination in yeast: alteration by multiple heterozygosities. Science **237:** 1459–1465.

Borts, R. H., W. Y. Leung, W. Kramer, B. Kramer, M. Williamson *et al.*, 1990 Mismatch repair-induced meiotic recombination requires the *pms1* gene product. Genetics **124:** 573–584.

Brown, T. C., and J. Jiricny, 1988 Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells. Cell **54:** 705–711.

Capecchi, M. S., 1980 High efficiency transformation by direct microinjection of DNA into cultured mammalian cells. Cell **22:** 479–488.

de Wind, N., M. Dekker, A. Berns, M. Radman and H. Te Riele, 1995 Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. Cell **82:** 321–330.

Dohet, C., S. Dzidie, R. Wagner and M. Radman, 1987 Large non-homology in heteroduplex DNA is processed differently than single base pair mismatches. Mol. Gen. Genet. **206:** 181–184.

Harris, S., K. S. Rudnicki and J. E. Haber, 1993 Gene conversions and crossing over during homologous and homeologous ectopic recombination in *Saccharomyces cerevisiae.* Genetics **135:** 5–16.

Kitts, P. A., and H. A. Nash, 1987 Homology-dependent interactions in phage lambda site-specific recombination. Nature **329:** 346–348.

Letsou, A., and R. M. Liskay, 1987 Effect of the molecular nature of mutation on the efficiency of intrachromosomal gene conversion in mouse cells. Genetics **117:** 759–769.

Liskay, R. M., and J. L. Stachelek, 1986 Information transfer between duplicated chromosomal sequences in mammalian cells involves contiguous regions of DNA. Proc. Natl. Acad. Sci. USA **83:** 1802–1806.

Liskay, R. M., J. L. Stachelek and A. Letsou, 1984 Homologous recombination between repeated chromosomal sequences in mouse cells. Cold Spring Harbor Symp. Quant. Biol. **49:** 183–189.

Liskay, R. M., A. Letsou and J. L. Stachelek, 1987 Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. Genetics **115:** 161–167.

Lukacsovich, T., D. Yang and A. S. Waldman, 1994 Repair of a specific double-strand break generated within a mammalian chromosome by yeast endonuclease I-*Sce*I. Nucleic Acids Res. **22:** 5649–5657.

Luria, S. E., and M. Delbrück, 1943 Mutations of bacteria from virus sensitivity to virus resistance. Genetics **28:** 491–511.

Modrich, P., and R. Lahue, 1996 Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annu. Rev. Biochem. **65:** 101–133.

Negritto, M. T., X. Wu, T. Kuo, S. Chu and A. M. Bailis, 1997 Influence of sequence identity on efficiency of targeted gene replacement. Mol. Cell. Biol. **17:** 278–286.

Parker, B. O., and M. G. Marinus, 1992 Repair of DNA heteroduplexes containing small heterologous sequences in *Escherichia coli.* Proc. Natl. Acad. Sci. USA **89:** 1730–1734.

Rayssiguier, C., D. S. Thaler and M. Radman, 1989 The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. Nature **342:** 396–401.

Rubnitz, J., and S. Subramani, 1984 The minimum amount of homology required for homologous recombination in mammalian cells. Mol. Cell. Biol. **4:** 2253–2258.

Shen, P., and H. V. Huang, 1986 Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. Genetics **112:** 441–457.

Shen, P., and H. V. Huang, 1989 Effect of base pair mismatches on recombination via the RecBCD pathway. Mol. Gen. Genet. **218:** 358–360.

Sia, E. A., R. J. Kokoska, M. Dominska, P. Greenwell and T. D. Petes, 1997 Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. Mol. Cell. Biol. **17:** 2851–2858.

Sokal, R. R., and F. J. Rohlf, 1981 *Biometry*, Ed. 2. Freeman and Company, San Francisco.

Southern, P. S., and P. Berg, 1982 Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early promoter. J. Mol. Appl. Genet. **1:** 327–341.

Štambuk, S., and M. Radman, 1998 Mechanism and control of interspecies recombination in *Escherichia coli.* I. Mismatch repair, methylation, recombination and replication functions. Genetics **150:** 533–542.

Taghian, D. G., H. Hough and J. A. Nickoloff, 1998 Biased short tract repair of palindromic loop mismatches in mammalian cells. Genetics **148:** 1257–1268.

Tran, H. T., D. A. Gordenin and M. A. Resnick, 1996 The prevention of repeat-associated deletions in *Saccharomyces cerevisiae* by mismatch repair depends on size and origin of deletions. Genetics **143:** 1579–1587.

Wagner, M. J., J. A. Sharp and W. C. Summers, 1981 Nucleotide sequence of the thymidine kinase gene of herpes simplex virus type 1. Proc. Natl. Acad. Sci. USA **78:** 1441–1445.

Waldman, A. S., and R. M. Liskay, 1987 Differential effects of base-pair mismatch on intrachromosomal versus extrachromosomal

recombination in mouse cells. Proc. Natl. Acad. Sci. USA **84:** 5340–5344.

Waldman, A. S., and R. M. Liskay, 1988 Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. Mol. Cell. Biol. **8:** 5350–5357.

Worth, L., S. Clark, M. Radman and P. Modrich, 1994 Mismatch repair proteins MutS and MutL inhibit RecA-catalyzed strand transfer between diverged DNAs. Proc. Natl. Acad. Sci. USA **91:** 3238–3241.

Yang, D., and A. S. Waldman, 1997 Fine-resolution analysis of products of intrachromosomal homeologous recombination in mammalian cells. Mol. Cell. Biol. **17:** 3614–3628.

Communicating editor: M. Lichten