MOLECULAR AND CELLULAR BIOLOGY, Dec. 1988, p. 5350–5357
0270-7306/88/125350-08$02.00/0

Vol. 8, No. 12

Dependence of Intrachromosomal Recombination in Mammalian Cells on Uninterrupted Homology

ALAN S. WALDMAN AND R. MICHAEL LISKAY*

Departments of Therapeutic Radiology and Human Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510

Received 29 April 1988/Accepted 9 September 1988

Recombination between a 360-base-pair (bp) segment of a wild-type thymidine kinase gene (*tk*) from each of three different strains (F, MP, and 101) of herpes simplex virus type one and a complete herpes simplex virus type 1 (strain F) *tk* gene containing an 8-bp insertion mutation was studied. The pairs of *tk* sequences resided as closely linked repeats within the genome of mouse LTK$^-$ cells. The frequency of recombination between sequences exhibiting 232 bp of uninterrupted homology and containing no mismatches other than the insertion mutation was comparable to the frequency of recombination between two sequences exhibiting four additional nucleotide mismatches distributed in such a way to preserve the 232-bp stretch of contiguous homology. In contrast, the placement of only two single-nucleotide mismatches (in addition to the insertion mutation) in such a manner to reduce the longest uninterrupted homology to 134 bp resulted in a 20-fold reduction in recombination. We conclude that the rate of intrachromosomal recombination in mammalian cells is determined by the amount of uninterrupted homology available and not by the total number of mismatches within a given interval of DNA. Furthermore, efficient recombination appears to require between 134 and 232 bp of uninterrupted homology; single-nucleotide heterologies are most likely sufficient to disrupt the minimal efficient recombination target. We also observed that if recombination was allowed to initiate within sequences exhibiting perfect homology, the event could propagate through and terminate within adjacent sequences exhibiting 19% base pair mismatch. We interpret this to mean that heterology exerts most of its impact on early rather than late steps of intrachromosomal recombination in mammalian cells.

Two approaches have been used to study the homology requirements of general recombination in vivo. One approach involves determining the rate of recombination as a function of the length of shared homology. In *Escherichia coli* the recombination rate is proportional to the length of homology, when this length exceeds approximately 30 to 50 base pairs (bp) (15, 23). Below 30 bp the rate of recombination drops off sharply. The interpretation of these data has been that the minimal target size for efficient *recBC*-mediated recombination in *E. coli* is about 30 bp (15). Recombination in bacteriophage T4 appears to have similar homology requirements (16). Experiments conducted in our laboratory have indicated that intrachromosomal recombination in mammalian cells drops off sharply when the amount of homology shared by two closely linked sequences is reduced from 295 to 200 bp (7). The results of one study of extrachromosomal recombination in mammalian cells are consistent with a minimal efficient target size of ~200 bp (13). (Low levels of recombination have been detected in some extrachromosomal recombination systems with sequences sharing less than 100 bp of homology [1, 4]). On the molecular level, the significance of recombination target sizes is not known, but the above studies have led to the idea that, in general, mammalian cells require substantially more homology than do bacteria for general recombination to proceed efficiently.

A second way of measuring the homology requirements is to determine the effect of base pair mismatch on the rate of recombination. In bacteria, 16% mismatch between sequences reduces phage-plasmid recombination rates by a factor of 100 (15). In addition, the rate of such recombination in *E. coli* is primarily determined by the lengths of perfect, uninterrupted stretches of homology rather than the overall percent homology between two sequences. The term minimal efficient processing segment (MEPS) has been used by Shen and Huang (15) to describe the minimal amount of uninterrupted homology needed for efficient recombination. The MEPS value for the *recBC*-mediated pathway of *E. coli* acting in phage-plasmid recombination is approximately 30 bp. Using this terminology, Shen and Huang (15) argue that the rate of recombination between any sequences is proportional to the total number of MEPS units shared by two sequences.

Previously, we examined the effect of 19% base pair mismatch on intrachromosomal recombination in mammalian cells using thymidine kinase gene (*tk*) sequences from herpes simplex virus types 1 and 2 (HSV-1 and HSV-2) (21). These studies indicated that intrachromosomal recombination can be reduced over 1,000-fold by 19% base pair mismatch. The mismatches between HSV-1 and HSV-2 *tk* genes are evenly distributed, with the longest stretch of perfect homology being 30 bp. In this work, using more closely matched sequences, we present experiments designed to determine whether the rate of intrachromosomal recombination in mammalian cells is governed primarily by the overall homology or, instead, by the length of uninterrupted homology exhibited by two sequences. In addition, these studies allowed us to make an assessment of which step(s) of recombination is most sensitive to heterology.

MATERIALS AND METHODS

Cell culture and transfections. Mouse L cells deficient in thymidine kinase (LTK$^-$ cells) were cultured and transfected by the calcium phosphate coprecipitate method or microinjection as previously described (21). Cell lines containing a

* Corresponding author.

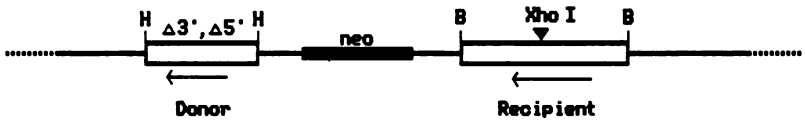

FIG. 1. Organization of the substrates used to study intrachromosomal recombination between defective *tk* sequences. All constructs are based on a pSV2neo-derived vector, indicated by solid lines, which encodes resistance to the neomycin analog G418. Each construct contains two defective HSV *tk* gene sequences oriented as direct repeats (open rectangles) and separated by about 4 kb of vector sequences. The arrows indicate the direction of transcription of the *tk* sequences in the native genes. One *tk* sequence, inserted at the *Hin*dIII (H) site of the vector, is a defective internal fragment of a *tk* gene with 5′ and 3′ deletions of the coding region. This *tk* fragment is intended to act as a donor of sequence information in gene conversion events. The other *tk* sequence, inserted at the *Bam*HI (B) site, is a complete HSV-1 F *tk* gene that has been mutated by an insertion of an 8-bp *Xho*I linker, as indicated. This defective gene is intended to act as a recipient in gene conversion events. Constructs were linearized by digestion with *Cla*I, and the linearized DNA was used to transfect mouse LTK$^-$ cells to produce cell lines containing one or two constructs stably integrated into the L-cell genome. Mouse cell genomic DNA is shown as broken lines flanking the construct.

low number of copies of the appropriate construct were identified as previously described (9).

Determination of intrachromosomal recombination frequencies. Fluctuation analyses were done as described previously (9). Ten independent subclones were used in the analysis of each cell line. Thymidine kinase-positive recombinants were selected in hypoxanthine-aminopterin-thymidine (HAT) medium (19).

Plasmid constructions. The vector used in all constructions was a derivative of pSV2neo (17), with restriction site alterations as described previously (8). *Xho*I linker insertion mutations of the HSV-1 (strain F) *tk* gene were gifts from D. Zipser and J. Kwoh. The *tk* genes from HSV-1 (strains 101 and MP) were gifts from W. C. Summers. The HSV-2 (strain 333) *tk* gene was a gift from D. Galloway. Mutant 8 contains an 8-bp *Xho*I linker inserted at nucleotide 1215 of the HSV-1 (strain F) *tk* gene, whereas mutant 28 contains a *Xho*I linker inserted at nucleotide 1036 of the gene (all numbering in this work is in accord with that of Wagner et al. [20]).

The mutant 28 *tk* gene (from strain F) was inserted on a 2.5-kilobase (kb) *Bam*HI fragment into the unique *Bam*HI site of the pSV2neo vector, after attachment of *Bam*HI linkers (New England BioLabs, Inc., Beverly, Mass.) to produce plasmid pTK28. The unique 361-bp *Nru*I-*Bal*I fragments from the wild-type *tk* genes from HSV-1 F, MP, and 101 were isolated and inserted (after attachment of *Hin*dIII linkers) into the unique *Hin*dIII site of plasmid pTK28 to produce, respectively, plasmids pF-F, pF-MP, and pF-101. These 361-bp fragments span the sequence from the *Nru*I site at nucleotide 805 to the *Bal*I site at nucleotide 1165 of the respective *tk* genes. These fragments do not encode functional *tk* genes owing to 5′ and 3′ deletions of the *tk* coding region.

A hybrid *tk* sequence consisting of sequence from the HSV-1 (strain F) *tk* gene and from the HSV-2 (strain 333) *tk* gene was constructed as follows. The fragment of the HSV-2 *tk* gene spanning the *Eco*RV site at nucleotide 848 to the *Pst*I site at nucleotide 1260 and the fragment of the HSV-1 *tk* gene spanning the *Pst*I site at nucleotide 1260 to the *Sma*I site at nucleotide 1621 were isolated. These two fragments were ligated together by their common *Pst*I ends. Importantly, the *Pst*I sites in the two genes are located at homologous positions (5, 18, 20), so the fragment produced by the ligation contains a contiguous *tk* coding region consisting of sequence from HSV-1 and HSV-2. *Hin*dIII linkers were attached to this hybrid fragment, and the fragment was inserted into the *Hin*dIII site of the pSV2neo vector. Into the unique *Bam*HI site was cloned a 2.5-kb fragment containing mutant 8 of the HSV-1 (strain F) *tk* gene to produce plasmid pHYB.

DNA sequence determination. The nucleotide sequence of the HSV-1 F *tk* gene between the *Nru*I site at nucleotide 805 and the *Bal*I site at nucleotide 1165 was determined by both the chemical degradation method of Maxam and Gilbert (11) and the synthesis method of Sanger et al. (14). Sequence information was independently determined for both strands. This sequence has been deposited with GenBank/EMBL (accession number JO3366).

DNA preparation and analysis. DNA was isolated and purified from tissue culture cells and subjected to Southern blotting analysis as previously described (8).

RESULTS

Efficient intrachromosomal recombination in mammalian cells requires uninterrupted homology. The experiments to be described involve studying recombination between HSV *tk* gene sequences that reside as closely linked direct repeats within the chromosomes of mouse LTK$^-$ cells. The *tk* sequences were transfected into the cells on constructs derived from pSV2neo (17). The experimental scheme is illustrated in Fig. 1. We were interested in determining whether the rate of recombination between closely linked chromosomal sequences is determined primarily by the overall homology or, instead, by the length of uninterrupted stretches of homology. To make such a determination, we derived cell lines containing construct pF-F, pF-MP, or pF-101 stably integrated into the host genome.

The important features of these recombination substrates are illustrated in Fig. 2A. Each construct contains an *Xho*I linker insertion mutant (mutant 28) of the HSV-1 strain F *tk* gene intended to serve as a recipient in gene conversions and a defective internal fragment of the wild-type *tk* gene from strain F, MP, or 101 to act as a donor in conversions. The length of the donor in each case is 360 bp. Importantly, we chose a donor of >300 bp so that the total length of homology of the two sequences would be greater than the minimum value established in our laboratory (7) as sufficient for efficient intrachromosomal recombination. By selecting for TK-positive segregants, we were able to monitor recombination events in which the insertion mutation in the recipient gene had been corrected by a transfer of wild-type sequence from the donor *tk* gene fragment. Only gene conversion events or double reciprocal exchanges were recoverable because single crossovers would result in 5′ or 3′ deletions of the *tk* gene. That the recoverable events actually represented nonreciprocal sequence exchanges (gene conversions) rather than double reciprocal exchanges was strongly suggested by recent studies from our laboratory (3).

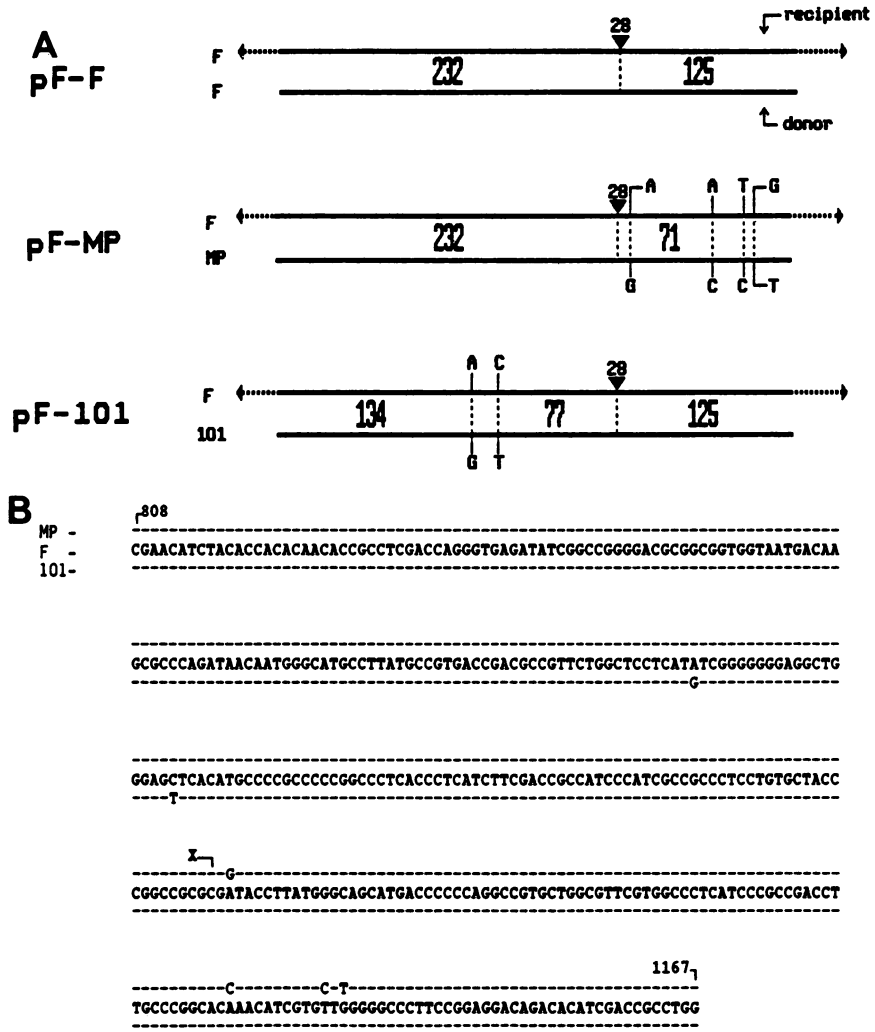

FIG. 2. Substrates pF-F, pF-MP, and pF-101. (A) Shown is a schematic illustration of the relevant *tk* sequences of each construct aligned in homologous register. In cell lines containing one of the above constructs, the *tk* sequences reside as closely linked direct repeats stably integrated into the L-cell genome, as illustrated in Fig. 1. In this figure, the direction of transcription of the *tk* sequences is from left to right. The position of *Xho*I linker insertion 28 is indicated by 28, and the nucleotide mismatches between the three pairs of *tk* sequences are also shown. Numbers shown indicate the lengths of the longer stretches of uninterrupted homologies in base pairs. The sequence alignment of pF-MP displays the greatest number of mismatches, while the pF-101 sequence alignment displays the least amount of contiguous homology. See text for details. (B) Donor nucleotide sequences from the *tk* genes from HSV-1 F, MP, and 101. The sequences illustrated are those of the internal fragments of the *tk* genes from strains F, MP, and 101 contained within constructs pF-F, pF-MP, and pF-101, respectively. Shown is the nucleotide sequence we determined for the strain F *tk* gene from nucleotide 808 to nucleotide 1167 (numbering according to Wagner et al. [20]). The position of the *Xho*I linker insertion in mutant 28 is indicated by the X (after nucleotide 1040). Differences between the strain F and strain MP *tk* sequences (12) are indicated above the strain F sequence, while differences between the strain F and strain 101 *tk* sequences (20) are indicated below the strain F sequence. Dashes indicate identity with the strain F sequence.

The nucleotide sequences of the internal donor fragments of the three *tk* genes used in constructs pF-F, pF-MP, and pF-101 are presented in Fig. 2B. When the sequences from strains F and MP were compared, four single-nucleotide mismatches were found. When the sequences from strains F and 101 were compared, two single-nucleotide mismatches were found.

The positions of the mismatches among the three *tk* sequences in relation to the position of the *Xho*I linker insertion mutation of mutant 28 are schematically illustrated in Fig. 2A. When the two strain F *tk* sequences present in construct pF-F were aligned in register, a stretch of 232 nucleotides of perfect sequence match and a stretch of 125 nucleotides of perfect sequence match resulted. These two regions of contiguous homology were separated by the heterology introduced by the 8-bp *Xho*I linker insertion in the mutant 28 gene. From previous studies (7), we inferred that the presence of the 232-bp stretch of homology should have provided sufficient homology to allow efficient recombination between the *tk* sequences. When the mutant strain F *tk* gene and the strain MP *tk* sequence present in pF-MP were aligned (Fig. 2A), the 232-bp stretch of perfect match present in pF-F was again present; however, owing to the four single-nucleotide mismatches between strains F and MP, the 125-bp stretch of perfect sequence homology of pF-F was interrupted. When the mutant strain F *tk* gene and the strain 101 *tk* sequence present in pF-101 were aligned (Fig. 2A), the 125-bp stretch of perfect homology present in

TABLE 1. Recombination frequencies of interstrain crosses of HSV-1 *tk* gene sequences

Cross	Line	Copy no.[a]	Cells tested (10^6)[b]	Colonies[c]	Recombination frequency[d]
F × F	1	1	226	52	2.3×10^{-7}
	2	2	250	53	1.0×10^{-7}
	3	2	270	73	1.3×10^{-7}
Avg					1.5×10^{-7}
F × MP	1	2	250	46	1.0×10^{-7}
	2	2	291	80	1.4×10^{-7}
	3	2	294	33	0.6×10^{-7}
Avg					1.0×10^{-7}
F × 101	1	1	280	0	$<4.0 \times 10^{-9}$
	2	1	280	1	4.0×10^{-9}
	3	2	260	9	1.7×10^{-8}
Avg					8.0×10^{-9}

[a] Number of copies of the relevant construct that are stably integrated into the genome of each line.

[b] Ten subclones were independently tested for each cell line. Shown is the total number of cells tested among all the subclones for each line.

[c] Number of HAT^r colonies.

[d] Determined for each cell line by dividing the number of HAT^r colonies by the number of cells tested in HAT medium and then dividing that quotient by the copy number of the line.

pF-F was present. However, the two single-base mismatches between strains F and 101 interrupted the 232-bp stretch of perfect match present in the pF-F and pF-MP substrates. The longest uninterrupted stretch of homology for pF-101 was only 134 bp.

Recombination frequencies were determined for cell lines containing construct pF-F, pF-MP, or pF-101 stably integrated into the L-cell genome. We reasoned that if the frequency of recombination between closely linked sequences within a mammalian chromosome is determined primarily by the overall percent homology of the sequences, then cell lines containing pF-MP should exhibit the lowest recombination frequency, since the *tk* sequences in this substrate contain the greatest degree of mismatch. If, on the other hand, the level of intrachromosomal recombination is determined primarily by the availability of contiguous, perfect sequence match, cell lines containing pF-101 might exhibit the lowest recombination frequency, since the *tk* sequences of this construct provide the shortest stretch of perfect match (134 bp; Fig. 2A). Furthermore, since previous studies conducted in our laboratory (7) had indicated that recombination drops about 10-fold when the amount of homology shared by two sequences is reduced from 295 to 200 bp, we realized that the 134 bp of uninterrupted homology provided by pF-101 might be below the threshold amount of homology needed to allow efficient recombination, if rates are indeed governed by the amount of contiguous homology available. In contrast, the pF-F and the pF-MP substrates should exhibit similar higher levels of recombination because both possess 232-bp stretches of perfect homology.

The results of fluctuation analyses performed on cell lines containing the pF-F, pF-MP, or pF-101 substrate are shown in Table 1. Cell lines containing pF-F or pF-MP gave comparable recombination frequencies ($\sim 10^{-7}$), whereas cell lines containing pF-101 exhibited frequencies approximately 20-fold reduced compared with frequencies determined for the control substrate pF-F (0.8×10^{-8}). We interpret these results to mean that the frequency of recombination is determined primarily by the lengths of uninterrupted stretches of homology rather than by the total number of mismatches within a given interval of DNA and that the *tk* sequences of pF-101 do not provide sufficient uninterrupted homology to allow efficient recombination. Southern hybridization analysis (Fig. 3) of several HAT^r colonies from each parental cell line confirmed that these cells did contain reconstructed *tk* genes produced by recombination events that eliminated the *Xho*I linker mutation.

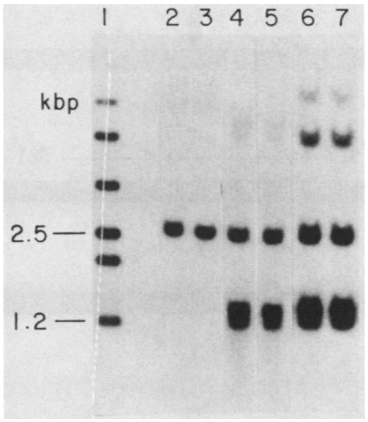

FIG. 3. Southern analysis of representative recombinant *tk* genes produced in lines containing pF-F, pF-MP, or pF-101. All parental cell lines contain recipient HSV-1 *tk* genes that contain *Xho*I linker insertion mutations and reside on 2.5-kb *Bam*HI fragments. Prior to recombination, the 2.5-kb *Bam*HI fragments in all lines can be cleaved by *Xho*I into a 1.3- and 1.2-kb fragment (not shown). Samples of DNA (8 μg) isolated from HAT^r recombinants were digested with *Bam*HI and *Xho*I and analyzed by Southern blotting with a probe specific for HSV-1 *tk* sequences. Lane 1, Molecular weight standards. Lanes 2 and 3, Recombinants derived from a cell line containing a single integrated copy of pF-F. These lines each display a 2.5-kb *Bam*HI fragment diagnostic for a reconstructed *tk* gene that no longer contains the *Xho*I linker insertion mutation and is therefore resistant to cleavage by *Xho*I. Lanes 4 and 5, Recombinants derived from a line containing two copies of pF-MP. These recombinants each display a 2.5-kb *Bam*HI fragment that contains a reconstructed *tk* gene and is therefore resistant to *Xho*I cleavage. These samples also display a second *Bam*HI fragment that contains a *tk* gene that remains uncorrected and is therefore cleaved by *Xho*I into a 1.3- and 1.2-kb fragment. Lanes 6 and 7, Recombinants derived from a cell line containing two copies of pF-101. The analysis is similar to that described for the samples displayed in lanes 4 and 5. Additional fragments of slower mobility visible in each of the lanes are junction fragments composed of vector sequences, the donor *tk* sequence inserted at the *Hin*dIII site of the vector (Fig. 1), and cellular DNA flanking the integrated construct. (Weak hybridization signals for some of these junction fragments are probably due to inefficient transfer of large DNA fragments to the filter as well as the small size of the sequence within these fragments that has homology to the probe.) Other than correction of the *Xho*I linker insertion mutations, no rearrangements of integrated constructs are detectable.

An early step of intrachromosomal recombination in mammalian cells is most sensitive to base pair mismatch heterology. The above experiments demonstrated that intrachromosomal recombination is sensitive to small amounts of heterology and suggested that recombination rates are primarily governed by the availability of contiguous uninterrupted homology. The experiments did not, however, address the issue as to which step(s) of recombination is most sensitive to heterology. It was possible that an early step (initiation) or

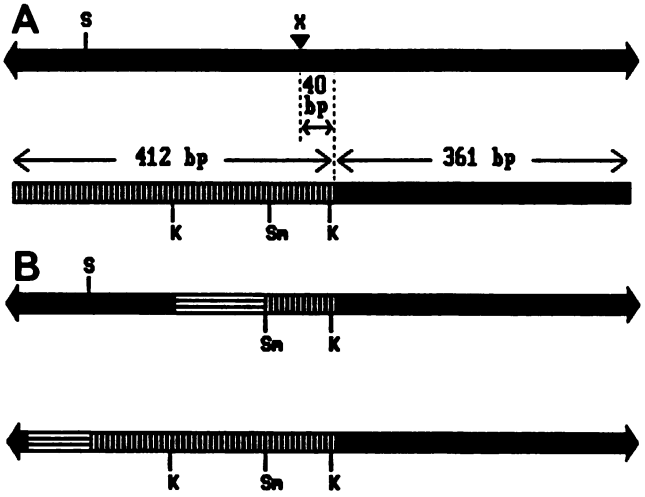

FIG. 4. (A) *tk* sequences contained in the hybrid donor substrate pHYB. Construct pHYB contains *Xho*I linker insertion mutant 8 of the HSV-1 (strain F) *tk* gene, acting as the recipient in gene conversions and represented by the upper sequence. The position of the *Xho*I linker insertion is indicated by X. The lower sequence, serving as the donor in gene conversions, is a hybrid sequence consisting of both HSV-1 (strain F) sequence (▬) and HSV-2 (strain 333) *tk* sequence (▥). The direction of transcription in this figure is from left to right. When the recipient mutant 8 gene is aligned in register with the hybrid donor *tk* fragment as illustrated, the *Xho*I linker insertion is opposite HSV-2 sequence, 40 bp upstream from the junction between HSV-1 and HSV-2 sequence in the donor. Within this 40-bp sequence, the HSV-1 and HSV-2 genes display nine single-nucleotide mismatches, with the longest stretch of uninterrupted homology being 11 bp (not shown). (B) Reconstructed *tk* genes produced by recombination between the *tk* sequences contained in pHYB. DNA was isolated from several HAT^r colonies arising from cell lines containing pHYB and analyzed by Southern blotting for the presence or absence of diagnostic *Xho*I (X), *Kpn*I (K), *Sac*I (S), and *Sma*I (Sm) sites. The analysis of two such genes is presented in Fig. 5. Shown are the deduced structures of the recombinant genes. Three of the genes analyzed are depicted by the upper map, while a fourth recombinant gene is depicted by the lower map. ▬, HSV-1 sequence; ▥, HSV-2 sequence; ▤, sequence that could not unequivocally be assigned as HSV-1 or HSV-2.

a later step (propagation, termination) or both were sensitive to the two mismatches present in pF-101. To make an assessment of which step(s) is most sensitive to heterology, we derived cell lines that contained construct pHYB (Fig. 4) stably integrated into their genomes. When the mutant 8 *tk* gene recipient sequence (from HSV-1 F) is aligned in register with the hybrid donor *tk* sequence (composed of HSV-1 and HSV-2 *tk* sequences) (Fig. 4), the *Xho*I linker insertion of mutant 8 is positioned opposite HSV-2 *tk* sequence. We have previously shown (21) that intrachromosomal recombination between HSV-1 and HSV-2 *tk* sequences occurs at an undetectably low rate ($<10^{-9}$ events per generation) owing to the 19% mismatch between the two genes. (In these previous studies, it was demonstrated by extrachromosomal recombination experiments that hybrid *tk* genes containing both type 1 and type 2 sequences are indeed functional.) The hybrid *tk* donor sequence used here contains 360 bp of HSV-1 *tk* sequence that is perfectly homologous to the mutant 8 sequence (Fig. 4).

We reasoned that if an early step(s) of recombination is most sensitive to heterology, TK-positive recombinants should be easily recoverable from lines containing pHYB,

TABLE 2. Recombination in cell lines containing pHYB

tk cross studied	No. of cell lines tested	No. of gene duplications (10^6)[a]	No. of HAT^r colonies	Recombination frequency[b]
pHYB	3	1,400	41	3.0×10^{-8}
HSV-1 *tk* × HSV-1 *tk*[c]	3	1,266	178	1.5×10^{-7}
HSV-1 *tk* × HSV-2 *tk*[d]	5	2,660	0	$<4.0 \times 10^{-10}$

[a] Calculated by taking the sum: $\Sigma (n_i)(p_i)$, where n = the copy number of the i^{th} cell line and p = the number of cells of the i^{th} cell line tested in HAT medium.

[b] Number of HAT^r colonies divided by the number of gene duplications tested.

[c] Data are derived from Table 1, cross F × F.

[d] Data are derived from previously published work (21).

since recombination should initiate within the homologous regions of the *tk* sequences provided by the hybrid donor. The predicted frequency of recombination should be comparable to the frequency observed for lines containing pF-F or pF-MP ($\sim 10^{-7}$) if propagation and termination are not blocked by the mismatches. If propagation or termination or both (in addition to initiation) are also sensitive to heterologies, then recombination leading to the correction of mutant 8, which requires extension through and resolution within 19% sequence mismatch, should occur at a significantly lower frequency.

Table 2 shows the results of fluctuation tests in which the frequency of intrachromosomal recombination between the hybrid *tk* donor sequence and mutant 8 recipient was determined. The frequency of recombination seen with pHYB was only 5-fold lower than the frequency seen with pF-F and at least 100-fold greater than the frequency previously measured for intrachromosomal recombination between HSV-1 and HSV-2 *tk* sequences.

The structures of several recombinant *tk* genes from cell lines containing pHYB were determined by Southern hybridization analysis (Fig. 5) and are displayed in Fig. 4. Each recombinant gene clearly contained sequence information from both HSV-1 and HSV-2 *tk* genes; the region of mutant 8 containing the insertion mutation was replaced by HSV-2 *tk* sequence in each case. The apparent conversion tracts were not all of equal length. Several of the recombinant genes exhibited conversion tracts between 80 and 200 bp in length, while one of the recombinants had a conversion tract greater than 300 bp in length. In every case, the conversion tract had apparently proceeded through mismatches beyond the distance needed to correct the mutation. Each of the recombinants contained a novel recombinant junction between HSV-1 and HSV-2 sequences.

DISCUSSION

Our previous studies (21) of intrachromosomal recombination in mammalian cells indicated that 19% base pair mismatch between two sequences represented sufficient heterology to reduce recombination over 1,000-fold. Such sensitivity to mismatch may be explained in two distinct ways.

Intrachromosomal recombination in mammalian cells might be governed by the global homology shared by two DNA molecules that are being tested for recombination. The sequences may be recombined only if there exists a sufficient percentage of nucleotide match along the sequences. (It is clear that a certain minimal number of nucleotide matches is required for recombination to proceed efficiently, for even recombination between perfectly homologous sequences re-

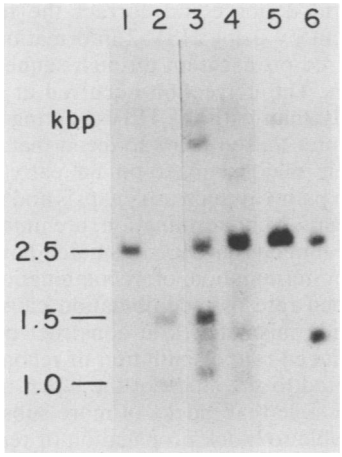

FIG. 5. Representative Southern analysis of two recombinants produced in cell lines containing pHYB. DNA samples (8 μg) isolated from HAT^r recombinants were digested with the indicated enzyme(s) and then analyzed by Southern blotting with a probe specific for HSV-1 *tk* sequences. Lanes 1 and 2 display DNA isolated from a single recombinant digested with *Bam*HI plus *Xho*I (lane 1) or *Bam*HI plus *Kpn*I (lane 2). This cell line contains a reconstructed gene as evidenced by the inability of the 2.5-kb *Bam*HI fragment to be cleaved by *Xho*I. The cleavage of the 2.5-kb *Bam*HI fragment into a 1.5- and a 1.0-kb fragment by *Kpn*I indicates that this recombinant gene acquired the 3′ *Kpn*I site but not the 5′ *Kpn*I site of the HSV-2 *tk* sequence of the hybrid donor (Fig. 4). Digestion with *Bam*HI plus *Sma*I (not shown) indicated that the *Sma*I site between the two *Kpn*I sites of the HSV-2 *tk* sequence was transferred to the reconstructed *tk* gene. Lanes 3 to 6 display DNA isolated from a different recombinant digested with *Bam*HI plus *Xho*I (lane 3), *Bam*HI plus *Kpn*I (lane 4), *Bam*HI alone (lane 5), or *Bam*HI plus *Sac*I (lane 6). This recombinant derives from a cell line containing two integrated copies of pHYB and so displays one corrected and one uncorrected *tk* gene as evidenced by one 2.5-kb *Bam*HI fragment that is resistant to *Xho*I digestion as well as one that is susceptible to *Xho*I digestion (*Xho*I cleaves the uncorrected gene into a 1.45- and a 1.05-kb fragment, lane 3). One of the *Bam*HI fragments is cleaved by *Kpn*I into 1.3-kb, 1.0-kb, and 200-bp fragments, the latter presumably running off the bottom of the gel (lane 4). This indicates that the recombinant *tk* gene acquired both of the *Kpn*I sites of the HSV-2 *tk* sequence contained in the hybrid donor (Fig. 4). Lane 6 shows that one of the 2.5-kb *Bam*HI fragments is cleaved by *Sac*I into two 1.25-kb fragments, while the other *Bam*HI fragment is resistant to *Sac*I cleavage. This indicates that the recombinant gene does not contain the *Sac*I site normally present in the HSV-1 gene but absent in the HSV-2 *tk* gene, suggesting that the recombination event had propagated beyond the position of the *Sac*I site, a distance of at least 300 bp. The structures of the two recombinant genes analyzed in this blot are schematically represented in Fig. 4B.

quires a minimum of about 200 bp of sequence match [7]). In short, this first model hypothesizes that each nucleotide of mismatch between two sequences would contribute equally to reducing the rate of recombination, with the prediction that for sufficiently long sequences (greater than 200 bp in length), recombination rates are governed by percent homology. Sensitivity to mismatch, in this model, would be synonymous with a requirement for a high percent homology.

In the second model, it is not the global percent mismatch per se that influences the recombination rate, and not every mismatch contributes equally to reducing the recombination rate. Rather, the manner in which mismatches are distributed is the factor of importance, and the lengths of uninterrupted homologies determine the recombination rate. Sensitivity to mismatch in this model would be synonymous with a requirement for a substantial region of perfect homology to serve as a suitable recombination target. This model invokes a localized search in which two sequences are scanned nucleotide by nucleotide in an attempt to locate a certain minimal length of uninterrupted, contiguous homology. This type of model was suggested by Shen and Huang (15) on the basis of their studies of bacteria.

The experiments presented in the first part of this report were designed to distinguish between the two general models described above. Our results are consistent with the second model, in which the recombination rate is a function of the amount of uninterrupted homology and which invokes a requirement for a certain minimal amount of contiguous homology for recombination to proceed efficiently. Although the *tk* sequences contained in pF-MP exhibited the greatest number of mismatches and hence the lowest overall percent homology of the three crosses studied, the frequencies determined for cell lines containing pF-MP were comparable to frequencies determined for cell lines containing pF-F, the construct with *tk* sequences exhibiting the least mismatch. The lowest frequencies of recombination were instead observed for cell lines containing pF-101, in which the *tk* sequences exhibited fewer mismatches than those of pF-MP but in which the longest uninterrupted stretch of homology was the shortest of the three *tk* crosses examined (Fig. 2). The only difference between the pF-F and pF-101 substrates is the presence of two single-nucleotide mismatches between the *tk* sequences in pF-101. That such a small sequence change had a dramatic effect on recombination demonstrates that recombination can be very sensitive to small heterologies; large amounts of heterology are not needed to perturb recombination.

Although we demonstrated that two single-nucleotide mismatches in pF-101 caused a significant decrease in recombination, it was a formal possibility that the effect was sequence specific and not general. For example, the two mismatches might have affected a crucial recombination sequence, a hotspot. We argue that the reduction in recombination was not due to such a sequence-specific effect but rather was due to a general heterology effect that resulted in an insufficient recombination target size, as follows. (i) Previously, we observed near equal recombination rates at five different sites within the HSV F *tk* gene (7, 9, 10), thereby suggesting that there are no recombination hotspots within the *tk* gene. (ii) The rate of recombination seen with sequences contained in construct pF-F in this study is in agreement with the rate measured in a previous study (7) involving a similarly sized fragment of the strain F *tk* gene mapping immediately adjacent to, but not overlapping, the interval of the *tk* gene studied here. This suggests that recombination between homologous sequences is a function of the size of a particular fragment of the *tk* gene rather than any special sequences. (iii) Our studies of recombination between a 1.2-kb fragment of the strain 101 *tk* gene and the mutant 28 strain F *tk* gene also argue against special sequences (9). In these experiments, the recombining strain 101 and strain F sequences shared over 250 bp of uninterrupted homology downstream from the *Xho*I linker insertion mutation. Although the very same two nucleotide mismatches present in the current study were also present in this previous work, the rate of recombination between the strain F gene and the 1.2-kb strain 101 sequence was equal to the rate of recombination between two comparable strain F sequences (9). Furthermore, the mismatches between the strain F and strain 101 sequences were frequently included in

the conversion tracts studied (9). This indicates that the effect of the two mismatches observed in the current work was completely overcome in crosses between longer sequences; the mismatches reduced the rate of recombination only when homology was limiting.

It was known from our previous studies (7) that, in crosses between homologous sequences, fragments serving as donors in gene conversion must be a minimum of about 200 bp in length for recombination to proceed efficiently, and the substrate pF-101 may be viewed with this in mind. The longest contiguous stretch of homology in pF-101 is 134 bp, and we measured a significantly reduced recombination rate for pF-101 relative to constructs pF-F and pF-MP, each of which contained perfect stretches of 232 bp. This suggests that 134 bp of contiguous homology represents an insufficient target for efficient recombination, whereas 232 bp of contiguous homology is a sufficient target. The issue as to whether recombination mechanisms can tolerate heterologies within the minimal recombination target is a question that has been raised previously (22). The experiments presented here suggest that the minimal target for efficient intrachromosomal recombination in mammalian cells must represent perfect homology; there is little tolerance for mismatch within the minimal recombination target. In fact, the borders of a suitable recombination target might be defined by single-nucleotide mismatches. With respect to the models discussed by Walsh (22) regarding how recombination rates might be affected by heterologies, our data suggest that gene conversion between duplicated sequences is incompletely described either by a simple *k*-hit model (in which recombination proceeds at a constant rate until *k* heterologies are introduced, at which point recombination is completely stopped) or by a more general model in which the conversion rate is a well-defined function of the fraction of sequence similarity. A modified combined model might be more appropriate, in which recombination is a function of blocks of perfect homology. The introduction of heterology might have little effect on the recombination rate until a sufficient number of mismatches are introduced and distributed so that the average distance between consecutive mismatches approaches the critical target size (MEPS). At this point, the recombination rate would drop rapidly with the introduction of additional mismatches.

Based on the results obtained in this study and the discussion above, the value of MEPS for mammalian intrachromosomal recombination may be approximated to be between 134 and 232 bp of perfect homology. This is substantially greater than the MEPS value of 30 bp determined for *recBC*-mediated recombination in *E. coli* (15, 23). A substantially larger MEPS value for mammalian cells might accommodate the larger genome size and the abundance of repeated sequences in mammalian DNA. An interesting issue is how a protein(s) can recognize one or two base pair mismatches within a 232 bp sequence. One might envision a mechanism involving cooperative interactions between several protein units, each individually responsible for a lesser amount of sequence homology.

An issue that we addressed is the question of which step(s) in recombination is most affected by base pair mismatch. Our previous results indicated that an HSV-2 donor sequence was unable to correct a mutation in an HSV-1 recipient sequence owing to 19% mismatch between the sequences. The rate of recombination was reduced >1,000-fold relative to a cross between two HSV-1 sequences (21). When we linked a 360-bp HSV-1 sequence to a donor HSV-2 sequence to promote initiation of recombination, we found that this hybrid donor could correct the mutation in the HSV-1 recipient by using HSV-2 information, even though to do so required propagation through sequences exhibiting 19% mismatch. The correction occurred at least 100 times more efficiently than with the HSV-2 acting alone as donor (21). We interpret these results to mean that small heterologies have their major impact on an early step(s) in the recombination pathway such as synapsis and that later steps, such as propagation and termination, are much less affected. In view of the finding that 19% mismatch has little effect on propagation or termination of recombination, we suggest that the reduced rate of recombination caused by the two single-nucleotide mismatches in construct pF-101 was due solely to a reduced rate of initiation of recombination. This work was limited to the study of the effect of small heterologies; it is possible that blocks of more substantial heterology might be able to block propagation or termination steps or both.

Of relevance to our findings that small heterologies appear to have their main impact on initiation of generalized recombination in mammalian cells, recent findings by Kitts and Nash (6) indicate that small heterologies are able to exert a significant effect on the branch migration process during site-specific recombination between lambda and *E. coli* attachment sites. It is possible that, in accord with the model proposed by Kitts and Nash (6), generalized recombination in mammalian cells initially requires a certain amount of branch migration to stabilize joint molecules following strand invasion. This initial stabilizing branch migration process, rather than synapsis or strand invasion, might be the actual step of recombination that is very sensitive to mismatch. After stabilization of the joint molecule is accomplished, propagation (e.g., further branch migration) and termination of the recombination event might proceed with little sensitivity to mismatch. Our experiments did not dissect the early steps of recombination so finely.

Examination of several reconstructed *tk* genes produced by recombination events involving the hybrid donor sequence indicated that once recombination was nucleated, the process could proceed through many mismatches and terminate. In one case, the conversion tract proceeded through 300 bp of sequences exhibiting 19% base pair mismatch. The results of this work therefore indicate that homologous sequences can promote the formation of recombinant junctions in nearby sequences that are otherwise sufficiently heterologous to preclude their recombination (21). A natural example of such an occurrence is illustrated by some products of unequal sister chromatid exchange events between closely linked alpha and delta glycophorin genes (2). Interestingly, although the two glycophorin genes share substantial regions of homology, the actual crossover point in at least one of the recombinants studied appears to reside within heterologous portions of the genes (2). We suggest that nearby regions of perfect homology permitted such events to occur.

As mentioned earlier, the constructs used for this study allow recovery of only gene conversions or double crossovers or both; single crossovers do not produce functional *tk* genes. However, simple gene conversion events are indistinguishable from double unequal sister chromatid exchanges by Southern blotting analysis. Although previous results from our laboratory strongly argue that the recombinants observed in these studies arose through gene conversions (3), it was nevertheless formally possible that the recombinants arose through double unequal sister chromatid exchanges. If this were indeed the case, then the hybrid

donor experiment could be reinterpreted to mean that the occurrence of a crossover within homologous sequences can stimulate an exchange in nearby, more heterologous sequences.

In conclusion, the results of this investigation suggest that mammalian intrachromosomal recombination involves a search for uninterrupted homology. Between 134 and 232 bp of uninterrupted homology are required for the efficient initiation of recombination, and initiation can be reduced by as few as two single-nucleotide mismatches. In contrast, propagation and termination appear to be substantially less sensitive to heterology such that once recombination is initiated in perfect homology the event can propagate and terminate in nearby, more heterologous sequences.

ACKNOWLEDGMENTS

We thank Jon Izant, Roni Bollag, and Alan Godwin for their helpful comments on the manuscript.

A.S.W. is a recipient of a Swebilius Cancer Research award, and R.M.L. is a Leukemia Society of America scholar. This work was supported by Public Health Service grants GM32741 to R.M.L. and PO1 CA39238 to William C. Summers from the National Institutes of Health.

LITERATURE CITED

1. **Ayares, D., L. Cherkuri, K.-Y. Song, and R. Kucherlapati.** 1986. Sequence homology requirements for intermolecular recombination in mammalian cells. Proc. Natl. Acad. Sci. USA **83**:5199–5203.
2. **Blumenfeld, O. O., A. J. Smith, and J. J. Moulds.** 1987. Membrane glycophorins of Dantu blood group erythrocytes. J. Biol. Chem. **262**:11864–11870.
3. **Bollag, R. J., and R. M. Liskay.** 1988. Conservative intrachromosomal recombination between inverted repeats in mouse cells: association between reciprocal exchange and gene conversion. Genetics **119**:161–169.
4. **Brenner, D. A., A. C. Smigocki, and R. D. Camerini-Otero.** 1985. Effect of insertions, deletions, and double-strand breaks on homologous recombination in mouse L cells. Mol. Cell. Biol. **5**:684–691.
5. **Kit, S., M. Kit, H. Qavi, D. Trkula, and H. Otsuka.** 1983. Nucleotide sequence of the herpes simplex virus type 2 (HSV-2) thymidine kinase gene and predicted amino acid sequence of thymidine kinase polypeptide and its comparison with the HSV-1 thymidine kinase gene. Biochim. Biophys. Acta **741**: 158–170.
6. **Kitts, P. A., and H. A. Nash.** 1987. Homology-dependent interactions in phage lambda site-specific recombination. Nature (London) **329**:346–348.
7. **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.
8. **Liskay, R. M., and J. L. Stachelek.** 1983. Evidence for intrachromosomal gene conversion in cultured mammalian cells. Cell **35**:157–165.
9. **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.
10. **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.
11. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
12. **McKnight, S. L.** 1980. The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene. Nucleic Acids Res. **8**:5949–5964.
13. **Rubnitz, J., and S. Subramani.** 1984. The minimum amount of homology required for homologous recombination in monkey cells. Mol. Cell. Biol. **4**:2253–2258.
14. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
15. **Shen, P., and H. V. Huang.** 1986. Homologous recombination in Escherichia coli: dependence on substrate length and homology. Genetics **112**:441–457.
16. **Singer, B. S., L. Gold, P. Gauss, and D. H. Dougherty.** 1982. Determination of the amount of homology required for recombination in bacteriophage T4. Cell **31**:25–33.
17. **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.
18. **Swain, M. A., and D. A. Galloway.** 1983. Nucleotide sequence of the herpes simplex virus type 2 thymidine kinase gene. J. Virol. **46**:1045–1050.
19. **Szybalski, W., E. H. Szybalska, and G. Ragni.** 1962. Genetic studies with human cell lines. Natl. Cancer Inst. Monogr. **7**:75–88.
20. **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.
21. **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.
22. **Walsh, J. B.** 1987. Sequence-dependent gene-conversion: can duplicated genes diverge fast enough to escape conversion? Genetics **117**:543–557.
23. **Watt, V. M., C. J. Ingles, M. S. Urdea, and W. J. Rutter.** 1985. Homology requirements for recombination in Escherichia coli. Proc. Natl. Acad. Sci. USA **82**:4768–4772.