Homology Dependence of Targeted Recombination at the Chinese Hamster *APRT* Locus

JULIA B. SCHEERER† AND GERALD M. ADAIR*

*The University of Texas M. D. Anderson Cancer Center, Science Park-Research Division,
Smithville, Texas 78957*

Received 3 May 1994/Returned for modification 20 June 1994/Accepted 8 July 1994

Using simple linear fragments of the Chinese hamster adenine phosphoribosyltransferase (*APRT*) gene as targeting vectors, we have investigated the homology dependence of targeted recombination at the endogenous *APRT* locus in Chinese hamster ovary (CHO) cells. We have examined the effects of varying either the overall length of targeting sequence homology or the length of 5' or 3' flanking homology on both the frequency of targeted homologous recombination and the types of recombination events that are obtained. We find an exponential (logarithmic) relationship between length of *APRT* targeting homology and the frequency of targeted recombination at the CHO *APRT* locus, with the frequency of targeted recombination dependent upon both the overall length of targeting homology and the length of homology flanking each side of the target gene deletion. Although most of the *APRT*⁺ recombinants analyzed reflect simple targeted replacement or conversion of the target gene deletion, a significant fraction appear to have arisen by target gene-templated extension and correction of the targeting fragment sequences. *APRT* fragments with limited targeting homology flanking one side of the target gene deletion yield proportionately fewer target gene conversion events and proportionately more templated extension and vector correction events than do fragments with more substantial flanking homology.

Gene targeting, targeted homologous recombination between an introduced DNA sequence and its endogenous chromosomal locus, is becoming increasingly useful as a means of genetic manipulation in mammalian cells as well as in yeast cells, permitting precise correction, disruption, replacement, or site-directed modification of virtually any gene or chromosomal locus for which a cloned sequence is available (8, 9, 27). However, much of our present understanding of the mechanisms of recombination is derived from studies in yeast cells; considerably less is known about recombination in mammalian cells. Most mammalian gene targeting studies have been directed toward disruption (knockout) or site-directed mutation of selected target gene loci in mouse embryo stem (ES) cells (8, 9), with their primary objective being to obtain the desired mutant mouse as quickly and easily as possible. Relatively few studies have examined targeted homologous recombination in cell types other than mouse ES cells.

We have previously described a model system that has proven to be quite useful for studying targeted homologous recombination at the endogenous adenine phosphoribosyltransferase (*APRT*) locus in Chinese hamster ovary (CHO) cells (1–3). Our system utilizes a hemizygous, nonreversible *APRT* deletion mutant (CHO-ATS49tg) as a recipient cell line, permitting direct selection of *APRT*⁺ recombinants arising by homologous recombination of *APRT* targeting sequences with the defective chromosomal gene locus. The 3-bp deletion in exon 5 of the CHO-ATS49tg *APRT* gene has eliminated an *Mbo*II restriction site, providing a useful restriction fragment length marker for Southern blot analysis. This cell line has been used for a variety of *APRT* targeting experiments, using

both replacement-type and insertion-type targeting vectors (1–3, 25, 40). In the present study, we have used simple, linear fragments of the CHO *APRT* gene (exclusive of any heterologous plasmid sequences) as targeting vectors to examine the relationship between length of targeting sequence homology and the frequency of targeted homologous recombination at the endogenous *APRT* locus. One of the distinct advantages of the CHO *APRT* system for such studies is that there are no *Alu*-equivalent repeats or other short (SINE) or long (LINE) interspersed repetitive elements within the 3.9-kb *Bam*HI fragment region containing the *APRT* target gene. Although the effects of such repetitive sequences on recombination in mammalian cells are presently unknown, they could present a number of complications for studies of homology dependence.

Studies of the homology dependence of targeted recombination at two other mammalian gene loci, the μ immunoglobulin gene in hybridoma cells (30) and the hypoxanthine phosphoribosyltransferase (*HPRT*) locus in mouse ES cells (9, 10, 14), present a somewhat confusing picture. Two of these studies demonstrate either a greater than linear (μ immunoglobulin [30]) or exponential (*HPRT* [10]) relationship between length of targeting sequence homology and targeted recombination frequency over the range of ~2 to 10 kb of homology. With a few exceptions (5, 7), most other studies of the homology dependence of recombination, including studies of recombination in T4 phage or *Escherichia coli* (29, 31, 41), recombination between transfected plasmid sequences in mammalian cells (7, 28), and mitotic recombination between direct or inverted chromosomal repeats in yeast (18, 33, 42) or mammalian (19) cells, have found a linear relationship between length of homology and recombination frequency, with a sharp dropoff in recombination efficiency below some minimum threshold length. Shen and Huang (29), who coined the term MEPS (minimal essential processing segment) to describe the threshold length of uninterrupted homology required for efficient recombination, argue that the rate of

* Corresponding author. Mailing address: The University of Texas M. D. Anderson Cancer Center, Science Park-Research Division, Smithville, TX 78957. Phone: (512) 237-2403. Fax: (512) 237-2437.

† Present address: Department of Molecular Genetics, The University of Texas M. D. Anderson Cancer Center, Houston, TX 77030.

recombination between any two sequences should be proportional to the number of MEPS units that it contains, and hence a linear relationship between length of homology and recombination frequency is expected. While the MEPS for eukaryotes appears to be somewhat larger than that for prokaryotes, the lengths of homology required for intrachromosomal recombination between direct repeats in yeast and mammalian cells are quite similar, approximately 250 to 300 bp (18, 19, 38, 39). Waldman and Liskay (39) have presented evidence that in mammalian cells, interruption of such contiguous stretches of homology by even single-nucleotide heterologies can dramatically reduce the frequency of intrachromosomal recombination. Recent studies comparing the effectiveness of isogenic versus nonisogenic targeting vector sequences for gene targeting in mouse ES cells have also demonstrated that even limited sequence divergence or scattered sequence heterogeneity can dramatically reduce gene targeting efficiency (10, 20, 35, 37), underscoring the importance of perfect, uninterrupted homology, and not just overall length of targeting homology, for effective targeted homologous recombination.

In this study, we have examined the effects of varying the overall length of targeting sequence homology, or specifically varying the length of 5' or 3' flanking homology, on the frequency of targeted homologous recombination at the endogenous CHO *APRT* locus and on the types of recombination events obtained. We have used 10 different linear fragments of the *APRT* gene, ranging in size from 896 to 3,275 bp, as targeting vectors to assess their relative efficiencies of targeted homologous recombination at the CHO *APRT* locus. A total of 169 independent *APRT*⁺ recombinants have been characterized by Southern blot analysis to determine the nature of the recombination events. We find that (i) there is a logarithmic relationship between the length of *APRT* targeting homology and the frequency of targeted recombination at the CHO-ATS49tg *APRT* locus, (ii) the frequency of targeted recombination at this locus is dependent upon both the overall length of targeting sequence homology and the length of homology flanking each side of the target gene deletion, (iii) a significant fraction of recombinants appear to arise by a novel recombinational pathway involving target gene-templated extension and correction of targeting fragment sequences, and (iv) targeting fragments with limited homology flanking one side of the target gene deletion yield proportionately fewer target gene conversion or gene replacement events and proportionately more targeting vector correction events than do fragments having more substantial flanking homology.

MATERIALS AND METHODS

Cell lines and culture conditions. The *APRT*[−] HPRT[−] CHO cell line used for these targeting experiments, CHO-ATS49tg, has been described previously (2, 3). This cell line, which was derived from CHO-AT3-2 (4), is hemizygous for the *APRT* locus, containing only a single, mutated copy of the *APRT* gene, with a 3-bp deletion in exon 5 that has eliminated an *Mbo*II restriction site (31a). No *APRT*⁺ revertants of CHO-ATS49tg have been obtained after plating more than 10⁹ cells into ALASA selection medium (25 μ M alanosine, 50 μ M azaserine, 100 μ M adenine) (2, 3, 25, 40). Cells were maintained as exponentially growing monolayer cultures in alpha-modified Eagle's minimal essential medium supplemented with 10% fetal bovine serum, penicillin, and streptomycin (2, 3).

***APRT* targeting fragments.** The 10 *APRT* gene fragments used as targeting vectors in these experiments (Fig. 1) were

FIG. 1. Maps of the CHO-ATS49tg *APRT* target gene locus, the *APRT* targeting fragments, and hybridization probes used for this study, showing relevant restriction endonuclease sites and fragment sizes. The vertical dashed line indicates the site of the 3-bp exon 5 *Mbo*II site deletion in the target gene. Shown for each *APRT* targeting fragment are its 5' and 3' terminal restriction sites, the overall length of targeting homology shared between the fragment and the target gene locus, and the lengths of 5' and 3' targeting homology flanking either side of the exon 5 *Mbo*II target gene deletion. Restriction endonuclease sites: B, *Bam*HI; Bn, *Ban*II; Bg, *Bgl*I; E, *Eco*RI; K, *Kpn*I; Mb, *Mbo*II; N, *Nde*I; Ps, *Pst*I; RV, *Eco*RV; X, *Xho*I; Xb, *Xba*I. Deleted restriction sites are shown in parentheses.

obtained by restriction endonuclease digestion of DNAs from plasmids containing subcloned fragments of the CHO *APRT* gene. Most of these fragments have been subcloned into pUC19 or pGEM vectors; others (e.g., *Bgl*I-*Bgl*I and *Eco*RI-*Nde*I) were isolated from plasmids containing slightly larger subcloned fragments (*Apa*LI-*Bam*HI and *Eco*RI-*Xba*I, respectively). For targeting experiments, the desired restriction fragments were isolated by agarose gel electrophoresis and purified with a Gene Clean kit (Bio 101) by a modified procedure (extended incubations at 4°C and repeated elutions). A full-

TABLE 1. Targeted recombination frequencies, using *APRT* gene targeting fragments

<i>APRT</i> targeting fragment ^a	Length of targeting homology (bp)	DNA concn (nM)	No. of cuvettes	Total no. of cells plated ^b	Total no. of <i>APRT</i> ⁺ colonies	Frequency of <i>APRT</i> ⁺ colonies		Relative targeting efficiency ^c
						Per 10 ⁷ cells	Per pmol of DNA ^d	
<i>Bam</i> HI- <i>Bam</i> HI ^e	3,881	3.9	2	1.2 × 10 ^{5f}	341	28,400	9,050	
(<i>Eco</i> RV)- <i>Bam</i> HI	3,275	4.6	12	2.6 × 10 ⁸	282	11.1	3.0	1:3,000
<i>Bgl</i> II- <i>Bgl</i> II	3,237	4.7	10	3.0 × 10 ⁸	472	16.0	4.3	1:2,100
<i>Eco</i> RI- <i>Bam</i> HI	2,573	5.9	4	2.6 × 10 ⁷	33	13.0	2.8	1:3,300
<i>Eco</i> RI- <i>Kpn</i> I	2,159	7.0	8	1.0 × 10 ⁸	41	4.1	0.73	1:12,600
<i>Xho</i> I- <i>Bam</i> HI	2,076	7.3	6	1.4 × 10 ⁸	21	1.5	0.26	1:35,000
Δ(<i>Xho</i> I)- <i>Bam</i> HI	~1,990	7.6	6	1.2 × 10 ⁸	9	0.8	0.13	1:71,000
<i>Pst</i> I- <i>Bam</i> HI	1,909	8.0	6	6.9 × 10 ⁷	5	0.7	0.12	1:75,000
<i>Eco</i> RI- <i>Xba</i> I	1,423	10.7	10	6.4 × 10 ⁷	16	2.5	0.23	1:39,000
<i>Eco</i> RI- <i>Nde</i> I	1,242	12.2	8	2.3 × 10 ⁸	11	0.5	0.05	1:185,000
<i>Ban</i> II- <i>Ban</i> II	896	17.0	8	2.1 × 10 ⁸	10	0.5	0.04	1:252,000

^a Deleted restriction sites are shown in parentheses.

^b Total number of cells plated into ALASA selection medium 48 h after electroporation.

^c Ratio of targeted recombination frequency to transfection frequency using the full-length *Bam*HI *APRT* gene fragment.

^d *APRT*⁺ colonies per 10⁷ cells plated 48 h after electroporation per picomole of targeting fragment DNA.

^e Transfection efficiency control; this fragment contains the entire functional *APRT* gene, including the promoter and poly(A) site.

^f Only a fraction of the cells transfected with the full-length *Bam*HI *APRT* gene fragment were plated.

length 3.9-kb *Bam*HI fragment containing the complete CHO *APRT* gene was also isolated. Purified *APRT* fragment DNAs were resuspended in 0.1× Tris-EDTA buffer at concentrations of 100 to 250 µg/ml.

Electroporations. Exponentially growing cells were harvested by trypsin-EDTA detachment and centrifugation and resuspended at 10⁷ cells per ml in sterile, ice-cold electroporation buffer (272 mM sucrose, 1 mM magnesium chloride, 7 mM sodium phosphate [pH 7.4]) containing *APRT* targeting fragment DNA (10 µg/ml). For each electroporation, 0.85 ml of DNA-cell suspension was transferred to a sterile electroporation cuvette (4-mm electrode gap; Bio-Rad), prechilled for 10 min on ice, and pulsed with 500 V at 25 µF, using a Bio-Rad Gene Pulser apparatus. After electroporation, the cells were maintained on ice for 10 min, diluted, and plated into fresh medium. A small portion of each electroporated cell population was plated into 60-mm-diameter dishes for plating efficiency and survival determinations. The rest of the cells were distributed into four to eight 100-mm-diameter dishes and maintained in alpha-modified Eagle's minimal essential medium for approximately 48 h before replating of each independent subpopulation into selection medium. The mean surviving cell fraction after electroporation in these experiments was 0.51 (range, 0.18 to 0.81), but we observed little correlation between survival level and recombination efficiency. To assure ourselves that observed differences in targeting frequency were not attributable simply to variations in electroporation efficiency, in some preliminary experiments cells were coelectroporated with *Xmn*I-linearized pMC1neoPolA plasmid DNA (1 µg/ml) along with *APRT* targeting fragment DNA (5 or 10 µg/ml), using the same protocol. As another monitor of electroporation efficiency, in several experiments cells were electroporated with a full-length 3.9-kb *Bam*HI fragment containing the complete *APRT* gene.

Selection of recombinants. Approximately 48 h after electroporation, the cells in each 100-mm-diameter dish were trypsinized, counted, and replated into ALASA selection medium to select for *APRT*⁺ recombinants (2, 3, 23). To ensure the independence of the *APRT*⁺ recombinants that were to be analyzed, only a single ALASA⁺ colony from each independent subpopulation of electroporated cells was picked for expansion, cryogenic preservation, and Southern blot analysis. In

preliminary experiments in which cells were coelectroporated with both pMC1neoPolA and *APRT* targeting fragment DNA, an aliquot of ~10⁵ electroporated cells was plated into G418 medium (400 µg/ml active portion) to monitor electroporation efficiency. Alanosine (NSC-529469) was obtained from the Drug Synthesis and Chemistry Branch of the National Cancer Institute.

DNA isolation and Southern blot hybridization. Genomic DNA was isolated from ALASA⁺ clones by using a simple salting-out procedure (21). Restriction digests of 10-µg samples of each genomic DNA were carried out with the reaction conditions recommended by suppliers, using 2 to 3 U of restriction enzyme per µg of DNA. After digestion with restriction endonucleases and electrophoresis (560 V-h) in 0.8% agarose gels, the DNAs were transferred to nitrocellulose membranes by capillary blotting (2, 4). The CHO *APRT* gene fragments that were used as hybridization probes for Southern blot analysis are shown in Fig. 1. Probe DNAs were labeled with [α-³²P]dCTP by random-primed labeling methods (Boehringer Mannheim) and separated from unincorporated nucleotides by using NICK columns (Pharmacia). Hybridizations, washings, and autoradiography were carried out as previously described (2, 4).

Statistical analysis. All statistical analyses were performed by using the Statview 4.0 software program (Abacus).

RESULTS

Relationship of targeted recombination frequency to the overall length of *APRT* targeting homology. To assess the homology dependence of targeted recombination at the CHO *APRT* locus, 10 different *APRT* targeting fragments (Fig. 1), ranging from 896 to 3,275 bp in length, were electroporated into CHO-ATS49tg cells, using the electroporation conditions and protocol described in Materials and Methods. None of these targeting fragments contain any *Alu*-equivalent repeats or other SINE or LINE elements. Table 1 summarizes the targeted recombination frequency data obtained from these experiments. *APRT*⁺ recombinants were obtained from even the smallest *APRT* targeting fragment tested, the 896-bp *Ban*II fragment. Absolute targeting frequencies have been normalized for fragment molarity so that the relative targeting

FIG. 2. Frequency of targeted recombination at the CHO *APRT* locus as a function of length of targeting homology. (A) Plot of targeted recombination frequency (*APRT*⁺ recombinants per 10^7 cells plated 48 h after electroporation per pmol of DNA) as a function of length of targeting homology for the 10 *APRT* targeting fragments shown in Fig. 1. Targeting frequency data are derived from Table 1. Each datum point represents the overall frequency of *APRT*⁺ recombinants obtained from targeting experiments using a given *APRT* targeting fragment (4 to 12 electroporations with each fragment). The curve was fitted by polynomial regression analysis ($R^2 = 0.875$; $P = 0.0007$). (B) Plot of the log frequency of *APRT*⁺ recombinants (per 10^7 cells plated 48 h after electroporation per pmol of DNA) as a function of length of targeting homology for the 10 *APRT* targeting fragments shown in Fig. 1. Targeting frequency data are derived from Table 1. Each datum point represents the overall frequency of *APRT*⁺ recombinants obtained from targeting experiments using a given *APRT* targeting fragment (4 to 12 electroporations with each fragment). The curve was fitted by polynomial regression analysis ($R^2 = 0.875$; $P = 0.0007$). (B) Plot of the log frequency of *APRT*⁺ recombinants (per 10^7 cells plated 48 h after electroporation per pmol of DNA) as a function of length of targeting homology, using the same data. The line was fitted by simple regression analysis of the data for all 10 *APRT* targeting fragments ($R^2 = 0.846$; $P = 0.0002$). The equation for this regression line is $y = 0.881x - 9.262$, where y represents the log frequency of targeted recombination at the CHO *APRT* locus and x represents the length of *APRT* targeting fragment homology. (C) Another plot of the same data, analyzed as two independent subsets, one (open circles) representing *APRT* fragments with at least 500 bp of targeting homology flanking each side of the exon 5 *Mbo*II site and the other (filled circles) representing *APRT* fragments having <500 bp of targeting homology on one side (either 5' or 3') of the *Mbo*II site. Lines were fitted by simple regression analyses of the data for each subset ($n = 4$ and $n = 6$, respectively).

efficiencies of these fragments can be compared on an equimolar basis. To monitor the effectiveness of our electroporation conditions and provide effectiveness of our electroporation controls for *APRT* targeting experiments, CHO-ATS49tg cells were electroporated with a 3,881-bp *Bam*HI fragment containing a complete, fully functional CHO *APRT* gene. With our standard electroporation conditions, the 3,881-bp *Bam*HI *APRT* fragment yielded *APRT*⁺ transfectants at a mean frequency of $\sim 2.8 \times 10^{-3}$, or $\sim 9,050$ transfectants per 10^7 cells plated 48 h after electroporation, per pmol of DNA (Table 1).

Comparison of the relative frequencies of *APRT*⁺ recombinants obtained in these gene targeting experiments (Table 1) makes two things evident. First, over the range of targeting homologies examined in this study (896 to 3,275 bp), targeted recombination is at least several orders of magnitude less efficient than nontargeted integration of DNA by mechanisms requiring little or no homology. Second, even relatively small increases in length of targeting homology can have profound effects on the efficiency of gene targeting. As shown in Table 1, a 3.6-fold increase in *APRT* targeting fragment homology (from 896 to 3,237 bp) resulted in an increase of 120-fold in the

frequency of targeted homologous recombination at the CHO-ATS49tg *APRT* locus.

Statistical analysis of the targeted recombination frequency data for the 10 *APRT* fragments in Table 1 indicated that there was a reasonably good correlation between targeting fragment length and recombination frequency ($R^2 = 0.758$; $P = 0.001$; correlation coefficient = 0.871). However, when these targeted recombination frequency data were plotted as a function of length of targeting homology (Fig. 2A), there clearly appeared to be a greater than linear dependence of recombination frequency on targeting fragment length. A reasonably good curve could be fitted to the data by polynomial regression analysis ($R^2 = 0.875$; $P = 0.0007$), suggesting a possible exponential relationship. When the same data were replotted as log frequency of *APRT*⁺ recombinants versus length of targeting homology (Fig. 2B), simple regression analysis gave an excellent straight-line fit ($R^2 = 0.846$; $P = 0.0002$), indicative of an exponential (logarithmic) relationship between length of *APRT* targeting homology and the frequency of targeted recombination at the CHO *APRT* locus.

Requirement for substantial targeting homology flanking each side of the target gene deletion. To assess the importance of the length of 5' or 3' flanking targeting homology on recombination frequency, the *APRT* targeting fragments shown in Fig. 1 were grouped into two subsets. One subset consisted of those *APRT* fragments having at least 500 bp of targeting homology flanking each side of the exon 5 *Mbo*II site, while the other subset represented fragments having less than 500 bp of targeting homology on one side (either 5' or 3') of the site. Targeted recombination frequency data for these two *APRT* fragment subclasses were then compared as separate subsets. When the data were plotted as log frequency of recombination versus targeting fragment length, and lines were fitted by simple regression analysis (as shown in Fig. 2C), the result was two noncontiguous regression lines with virtually identical slopes. The slope and y intercept of the regression line for fragments having at least 500 bp of targeting homology flanking each side of the exon 5 *Mbo*II site (open circles in Fig. 2C) were 0.538 and -8.160 , respectively, compared with 0.536 and -8.809 for fragments having less than 500 bp of targeting homology on one side (either 5' or 3') of the *Mbo*II site (filled circles in Fig. 2C). These results suggested that fragments with less than 500 bp of flanking *APRT* targeting homology on one side of the target gene deletion are less effective substrates for targeted recombination at the CHO *APRT* locus than similarly sized fragments that have at least 500 bp of targeting homology flanking each side of the deletion site, regardless of their overall length of targeting homology.

Dependence of targeted recombination frequency on the lengths of 5' flanking and 3' flanking targeting homology. Further analysis of targeted recombination data for two other subsets of *APRT* targeting fragments allowed us to examine the effects of specifically decreasing the length of either 5' or 3' flanking targeting homology while holding the length of homology on the other side constant. One set of fragments shared the same 3' terminus (the 3' *Bam*HI restriction site) and a constant length of 3' flanking targeting homology (1,597 bp) but varied in their lengths of 5' flanking homology (from 309 to 1,675 bp); the other set of fragments had the same 5' terminus (the *Eco*RI restriction site) and a constant length of 5' flanking targeting homology (973 bp) but varied in their lengths of 3' flanking homology (from 266 to 1,597 bp). When targeted recombination frequencies for the first set of fragments were plotted as log frequency of *APRT*⁺ recombinants versus length of 5' flanking targeting homology (Fig. 3A), a regression line with a slope of 1.118 and a y intercept of -8.156 was obtained ($R^2 = 0.815$; $P = 0.0358$). When targeted recombination frequencies for the second set of fragments were plotted as log frequency of *APRT*⁺ recombinants versus length of 3' flanking targeting homology (Fig. 3B), the resulting regression line ($R^2 = 0.904$; $P = 0.0493$) had approximately the same slope (1.124) and only a slightly lower y intercept (-8.374) than the regression line in Fig. 3A. These results suggested that the frequency of targeted recombination at the CHO-ATS49tg *APRT* locus increases as a function of increasing length of targeting homology flanking either side of the target gene deletion and that there appears to be a greater than linear relationship between the length of 5' or 3' flanking homology and the frequency of targeted recombination.

Further analysis of our targeting frequency data revealed that the frequency of targeted recombination at the CHO *APRT* locus appeared to be proportional to the product of the lengths of 5' and 3' flanking targeting homology. When targeted recombination frequencies for all 10 *APRT* fragments were plotted as frequency of *APRT*⁺ recombinants versus the product of the lengths of 5' flanking and 3' flanking targeting

FIG. 3. Frequency of targeted recombination at the CHO *APRT* locus as a function of length of 5' flanking or of 3' flanking targeting homology. (A) Plot of the log frequency of *APRT*⁺ recombinants (per 10^7 cells plated 48 h after electroporation per pmol of DNA) as a function of length of 5' flanking *APRT* targeting homology. These five fragments all have the same length (1,597 bp) of *APRT* targeting homology flanking the 3' side of the exon 5 *Mbo*II site but vary in their lengths of 5' flanking homology. The line was fitted by simple regression analysis. (B) Plot of the log frequency of *APRT*⁺ recombinants (per 10^7 cells plated 48 h after electroporation per pmol of DNA) as a function of length of 3' flanking *APRT* targeting homology. These four fragments all have the same length (973 bp) of *APRT* targeting homology flanking the 5' side of the exon 5 *Mbo*II site but vary in their lengths of 3' flanking homology. The line was fitted by simple regression analysis.

homology, simple linear regression analysis gave a straight-line fit with a slope of 1.606 and a y intercept of -0.566 ($R^2 = 0.890$; $P = <0.0001$).

Southern blot hybridization analysis of *APRT*⁺ recombinants. CHO-ATS49tg cells contain a single, mutationally altered copy of the *APRT* gene. A 3-bp *APRT* deletion has eliminated the exon 5 *Mbo*II restriction site in this cell line, providing a convenient restriction fragment length polymorphism (Fig. 4A, lane 2) that can be used to monitor targeted correction of the *APRT* target gene deletion. Southern blot analyses of *APRT*⁺ recombinants obtained from previous CHO *APRT* targeting experiments using conventional insertion-type or replacement-type *APRT* targeting vectors (2, 3, 6, 23, 25) have identified three general classes of recombinants, reflecting at least three different types of targeted homologous recombination events: (i) target gene correction or conversion, (ii) targeted insertion at the chromosomal *APRT* locus, or (iii) target gene-templated extension and correction of the *APRT* targeting sequence. These three recombinant classes are readily distinguishable on the basis of their distinctive and diagnostic restriction fragment patterns.

To determine the nature of the recombination events responsible for the generation of the *APRT*⁺ recombinants

FIG. 4. Southern blot hybridization analysis of wild-type CHO, CHO-ATS49tg, and representative *APRT*⁺ recombinant clone DNAs following digestion with *Mbo*II (A), *Bgl*II (B), *Eco*RI (C), *Bam*HI (D), *Pst*I (E), and *Xba*I (F). Lane 1, wild-type CHO cells; lane 2, CHO-ATS49tg cells; lane 3, clone EB-5, an *APRT*⁺ recombinant showing target gene conversion; lane 4, clone EB-10, an *APRT*⁺ recombinant showing targeted insertion of *APRT* fragment sequences at the target gene locus; lane 5, clone EB-4, an *APRT*⁺ recombinant showing templated extension and correction of the *APRT* targeting fragment sequence. *APRT* probes used for hybridization (see Fig. 1): (A) the 1.4-kb *Eco*RI-*Xba*I fragment; (B) the 1.3-kb *Bam*HI-*Eco*RI fragment; (C to F) the 3.9-kb *Bam*HI fragment.

obtained in the targeting experiments in Table 1, DNA samples from a total of 169 independent recombinant clones were digested with *Mbo*II and other restriction enzymes and subjected to Southern blot hybridization analysis, using the *APRT* probes shown in Fig. 1. Results of these analyses are summarized in Table 2. Restriction fragment patterns for *APRT*⁺ recombinants representing each of the three major classes are shown in Fig. 4.

Correction or conversion of the *APRT* target gene mutation. Seventy-seven of the 169 *APRT*⁺ recombinants analyzed showed simple correction or conversion of the CHO-ATS49tg target gene deletion (Table 2), with precise replacement of the three deleted bases and restoration of the exon 5 *Mbo*II restriction site. These recombinants appear to have arisen by targeted replacement or gene conversion. Such events would presumably require homologous interactions of an *APRT*

targeting fragment with sequences flanking both sides of the target gene deletion. Restoration of the wild-type (WT) sequence and exon 5 *Mbo*II restriction site results in loss of the 2.0-kb *Mbo*II fragment present in CHO-ATS49tg and reappearance of the 1.5-kb *Mbo*II fragment characteristic of the WT Chinese hamster *APRT* gene, as shown in Fig. 4A (lane 3). Target gene convertants should retain normal, WT *APRT* restriction fragment patterns for all other restriction enzymes (Fig. 4B to E, lanes 3), assuming that there are no other random integrations of targeting fragment sequences. Nontargeted integrations are readily identifiable by their lack of hybridization to 5'-end-specific *APRT* probes.

Targeted insertions of *APRT* targeting fragment sequences into the chromosomal locus. Linear replacement-type vectors such as the *APRT* targeting fragments were expected to be poor recombination substrates for targeted insertions (3, 23). Targeted insertion of such fragment sequences would presumably require either single-crossover, reciprocal recombination of a circularized *APRT* targeting fragment with the target gene locus or targeted replacement of the chromosomal *APRT* gene sequences by a linear or circularized head-to-tail concatamer of *APRT* targeting fragments (3, 14–16, 23). Nevertheless, 10 of the 169 *APRT*⁺ recombinants analyzed showed what appeared to be targeted insertions of *APRT* targeting fragment sequences at the CHO-ATS49tg chromosomal *APRT* locus.

Targeted insertions should result in the generation of a full-length, corrected *APRT* gene along with a partial duplication of *APRT* gene sequences at the original target gene locus. When genomic DNA from a targeted insertion recombinant is digested with restriction endonucleases that cleave at sites outside the targeting fragment sequence, one would expect to see a consistent increase in restriction fragment sizes compared with those for WT CHO or CHO-ATS49tg DNAs. Digestions with restriction enzymes cleaving at a unique site within the *APRT* targeting fragment should yield, in addition to the original restriction fragments, a new fragment, which should correspond in size to the length of the inserted sequence.

Restriction fragment patterns for one such targeted insertion clone, EB-10 (obtained from an experiment using the 2,573-bp *Eco*RI-*Bam*HI *APRT* fragment), are shown in lanes 4 of Fig. 4A to F. Digestion of EB-10 genomic DNA with *Bgl*II (which does not cut within the targeting fragment sequence) yielded a 6.8-kb fragment (Fig. 4B, lane 4) instead of the 4.7-kb fragment normally seen in WT CHO (lane 1) or CHO-ATS49tg DNAs (lane 2). Digestion of EB-10 DNA with either *Pst*I or *Xba*I (each of which has a unique cleavage site within the targeting fragment sequence) revealed a new 2.1-kbp fragment (Fig. 4E and F, lanes 4) in addition to the fragments normally seen following digestion of WT CHO (lanes 1) or CHO-ATS49tg DNAs (lanes 2) with these restriction enzymes. Thus, EB-10 appears to have an insertion of ~2.1 kb of *APRT* targeting sequences into the CHO-ATS49tg target gene locus. This insertion is ~500 bp shorter than would be expected based on the size of the *Eco*RI-*Bam*HI targeting fragment, suggesting that there must have been considerable exonucleolytic degradation of the ends of the targeting fragment molecules prior to end joining and formation of the circularized or concatamerized recombinational intermediates that were ultimately integrated into the CHO-ATS49tg target gene locus by reciprocal recombination or gene replacement events. Furthermore, restriction digests of EB-10 DNA with *Eco*RI (Fig. 4C, lane 4) or *Bam*HI (Fig. 4D, lane 4) each revealed an increase of ~2.1 kb in size of an existing restriction fragment, rather than the appearance of a new 2.1-kb restriction fragment, suggesting that both the 5' (*Eco*RI) and 3' (*Bam*HI) terminal

TABLE 2. Analysis of *APRT*⁺ recombinants obtained in CHO *APRT* gene targeting experiments

<i>APRT</i> targeting fragment ^a	Length of targeting homology (bp)	No. of <i>APRT</i> ⁺ clones analyzed	No. of target gene convertants	No. of targeted insertions	No. of targeting fragment corrections	No. with correction of both target gene and targeting fragment
(<i>EcoRV</i>)- <i>Bam</i> HI	3,275	31	16	2	13	0
<i>Bgl</i> II- <i>Bgl</i> II	3,237	30	15	1	14	0
<i>Eco</i> RI- <i>Bam</i> HI	2,573	31	19	2	9	1
<i>Eco</i> RI- <i>Kpn</i> I	2,159	36	18	2	16	0
Totals for fragments with >500 bp 5' and 3' flanking homologies ^b		128	68	7	52	1
% in each class			53	5	41	1
<i>Xho</i> I- <i>Bam</i> HI	2,076	8	3	0	5	0
Δ (<i>Xho</i> I)- <i>Bam</i> HI	~1,990	5	0	0	5	0
<i>Pst</i> I- <i>Bam</i> HI	1,909	4	3	1	0	0
<i>Eco</i> RI- <i>Xba</i> I	1,423	13	2	1	10	0
<i>Eco</i> RI- <i>Nde</i> I	1,242	5	0	0	5	0
<i>Ban</i> II- <i>Ban</i> II	896	6	1	1	3	1
Totals for fragments with <500 bp 5' or 3' flanking homology ^b		41	9	3	28	1
% in each class			22	7	68	2
Totals for all <i>APRT</i> targeting fragments		169	77	10	80	2
% in each class			46	6	47	1

^a Deleted restriction sites are shown in parentheses.

^b Contingency chi-square analysis indicated that the observed differences in recombinant-type distributions for the two targeting fragment subclasses are statistically significant ($\chi^2 = 12.546$; $P = 0.0057$; contingency coefficient = 0.263).

restriction sites have been lost from the inserted targeting fragment sequences.

Southern blot analyses of the other nine targeted insertion clones suggested that in the majority of these recombinants, the inserted sequence appeared to be ~50 to ~900 bp shorter than predicted based on the size of the *APRT* targeting fragment. Furthermore, in all 10 targeted insertion clones, the 5' and 3' terminal restriction sites have been lost from the inserted targeting fragment sequences (data not shown). A more detailed analysis of the structures of these recombinants and the nature of the recombinational events responsible for their generation will be published elsewhere.

Templated extension and correction of the targeting fragment sequences. Almost half of the recombinants analyzed (80 of 169) represent what we have previously termed class III or vector correction recombinants (2, 3, 23). *APRT*⁺ recombinants of this type retain an uncorrected *APRT* target gene at the original chromosomal locus but also contain a second, randomly integrated, WT copy of the *APRT* gene, apparently derived from a corrected targeting fragment sequence. Since none of the *APRT* targeting fragments used in this study contain any upstream flanking homology that would permit gap repair or replacement of missing promoter and 5' *APRT* coding sequences by conventional recombination pathways, such targeting fragment correction events are presumed to arise by a novel recombinational pathway involving target gene-templated extension and correction of the 5'-truncated targeting fragment *APRT* sequence, followed by its release and random (nonhomologous) integration at another site in the genome.

Restriction fragment patterns for one of these targeting fragment correction recombinants, clone EB-4 (derived from an experiment using the 2,573-bp *Eco*RI-*Bam*HI *APRT* fragment as a targeting vector), are shown in lanes 5 of Fig. 4A to E. *Mbo*II digestion of EB-4 genomic DNA (Fig. 4A, lane 5)

revealed retention of the 2.0-kb *Mbo*II *APRT* gene fragment characteristic of CHO-ATS49tg (lane 2) and reappearance of the 1.5-kb *Mbo*II fragment characteristic of the WT CHO *APRT* gene (lane 1). When genomic DNAs from either target gene convertants or targeted insertion recombinants are digested with restriction enzymes that cut outside of the *APRT* target gene and hybridized with a 5'-end-specific *APRT* probe, they should show only a single restriction fragment for each enzyme, characteristic of the target gene locus. Targeting fragment correction recombinants, however, would be expected to show two restriction fragments: one fragment representing the target gene locus and a second fragment of variable size, created by ectopic integration of a corrected *APRT* targeting fragment sequence. As shown in Fig. 4B (lane 5), hybridization of *Bgl*II-digested EB-4 DNA with the 1.3 kb *Bam*HI-*Eco*RI probe reveals two fragments: a 4.7-kb *Bgl*II fragment characteristic of the targeted chromosomal *APRT* locus and a much larger fragment representing a corrected *APRT* targeting fragment sequence.

Templated extension of the *APRT* targeting fragment can apparently proceed for considerable distances beyond the 5' end of the *APRT* gene. In most of the fragment correction recombinants analyzed, the *APRT* targeting fragment sequences have been extended at least far enough to acquire an *Eco*RI site which is ~7.1 kb upstream of the internal *Eco*RI site in the *APRT* gene, as well as a *Pst*I site which is ~3.8 kb upstream of the internal *Pst*I site. This can be seen for the targeting fragment correction recombinant EB-4 in Fig. 4. Greater relative intensities of hybridization are seen for both the 7.1-kb *Eco*RI fragment (Fig. 4C, lane 5) and the 3.8-kb *Pst*I fragment (Fig. 4E, lane 5) of EB-4 than for the same fragments from wild-type CHO (lane 1), CHO-ATS49tg (lane 2), or the targeted convertant EB-5 (lane 3), suggesting duplication of these genomic sequences by templated extension of the targeting fragment in EB-4. Similar results were obtained with the

1.3-kb *Bam*HI-*Eco*RI probe, which specifically hybridized to these two fragments (results not shown). In contrast to the restriction fragment patterns obtained with *Eco*RI and *Pst*I, *Bam*HI digestion of EB-4 (Fig. 4D, lane 5) yielded a 3.9-kb fragment characteristic of the *APRT* target gene locus, plus a much larger fragment, presumably representing the ectopically integrated, extension-corrected targeting fragment sequence. Thus, while templated extension of the *Eco*RI-*Bam*HI targeting fragment appears to have proceeded for at least 7.1 kb in the 5' (upstream) direction in EB-4, the targeting fragment was apparently not extended at all in the 3' (downstream) direction, as that should have regenerated the 3' *Bam*HI site, resulting in a 3.9-kb *Bam*HI fragment for both the extension-corrected targeting sequence and the *APRT* target gene.

Effects of limited 5' or 3' targeting homology on the types of targeted recombination events obtained in *APRT* targeting experiments. The relatively poor targeting efficiencies of some of these fragments made it difficult to isolate and analyze enough recombinants to permit direct fragment-to-fragment comparisons. Therefore, to assess the effects of limited 5' or 3' flanking targeting homology on the nature of homologous recombination events that were obtained in these *APRT* targeting experiments, we compared the recombinant class distributions for the same two subsets of targeting fragments that had been previously used for recombination frequency comparisons: *APRT* fragments that had at least 500 bp of targeting homology flanking each side of the exon 5 *Mbo*II site, versus fragments that had less than 500 bp of flanking homology on one side. As shown in Table 2, comparison of the relative frequencies of the various recombinant classes that were obtained with these two subsets of targeting fragments revealed a rather dramatic shift in the types of recombinants produced as the length of 5' or 3' flanking targeting homology became limited. Contingency chi-square tests indicated that the observed differences in recombinant-type distributions for these two targeting fragment subclasses were statistically significant ($\chi^2 = 12.546$; $P = 0.0057$). With *APRT* targeting fragments having at least 500 bp of targeting homology flanking each side of the exon 5 *Mbo*II site, the majority of *APRT*⁺ recombinants arose by target gene conversion (~53%); only about 40% of the recombinants from such experiments arose by templated extension and correction of the targeting fragment sequence. In contrast, targeting fragments with less than 500 bp of flanking homology on one side of the *Mbo*II site appeared to produce *APRT*⁺ recombinants primarily by templated extension and correction of the targeting fragment sequence (nearly 70% of all recombinants), with only about 20% of the recombinants reflecting target gene conversion events. Targeted insertions represented only a small fraction (~5 to 7%) of the recombinants in both sets of data. The *Eco*RI-*Nde*I fragment (Fig. 1), which has only 266 bp of 3' targeting homology and extends less than 20 bp beyond the normal *APRT* polyadenylation sequence, yielded only targeting sequence correction events (Table 2).

DISCUSSION

Using simple linear fragments of the Chinese hamster *APRT* gene as targeting vectors, we have examined the effects of varying either the overall length of targeting sequence homology or the length of 5' or 3' flanking homology on the frequency of targeted recombination at the CHO *APRT* locus and on the types of recombination events that are obtained. Comparison of the relative targeting efficiencies of 10 different *APRT* fragments, ranging in size from 896 to 3,275 bp, revealed a striking, logarithmic relationship between the overall length

of targeting sequence homology and the frequency of targeted recombination at the CHO-ATS49tg *APRT* locus. This is the second mammalian gene locus to show such a logarithmic or exponential dependence of targeted recombination frequency on the length of targeting homology. Our results are remarkably consistent with those obtained by Deng and Capecchi (9, 10) in their studies of the homology dependence of targeted recombination at the *HPRT* locus in mouse ES cells. Both of these gene loci appear to show a stronger dependence of targeted recombination on length of targeting sequence homology than has been reported for the μ immunoglobulin gene in hybridoma cells (30).

Our results indicate that the frequency of targeted recombination at the CHO-ATS49tg *APRT* locus is dependent upon both the overall length of targeting sequence homology and the length of targeting homology flanking each side of the target gene deletion. A number of targeting studies have reported that sequence replacement vectors with less than ~1 kb of targeting homology on one arm are generally poor substrates for targeted gene replacement, showing markedly lower targeting efficiencies and yielding very few simple replacement events (14–17, 30, 36). In *HPRT* targeting experiments in mouse ES cells, Hasty et al. (16) found that a sequence replacement vector that contained ~6.1 kb of targeting homology but only 472 bp of 3' flanking homology yielded only about one-third as many *HPRT*⁺ recombinants as a vector that contained approximately the same overall length of targeting homology but more (1.2 kbp) 3' flanking homology. They failed to recover any recombinants from *HPRT* targeting experiments using replacement vectors having less than 761 bp of 5' flanking targeting sequence homology. Similarly, we found that fragments having less than 500 bp of flanking *APRT* targeting homology on one side of the target gene deletion are much less effective substrates for targeted recombination than similarly sized fragments with at least 500 bp of targeting homology flanking each side of the deletion site, regardless of their overall length of targeting homology. Our experiments indicated that the frequency of targeted recombination at the CHO *APRT* locus increases logarithmically with increasing length of flanking homology on either the 5' or 3' side of the target gene deletion.

Our findings are also consistent with those of Jiang and coworkers (17), who found that the frequency of marker rescue by targeted recombination of replacement fragments with the μ immunoglobulin gene in mutant hybridoma cell lines was much lower when the marker was within ~400 bp of the end of the targeting homology than when the marker was near the middle. They argued that if marker rescue depends on the occurrence of two crossovers (one on each side of the mutation being corrected) and the probability of a crossover in each region is proportional to the segment length, then the frequency of marker rescue should be proportional to the product of the lengths of targeting homology flanking either side of the mutation being corrected. Their results appeared to fit this hypothesis. We found that the frequency of targeted recombination at the CHO *APRT* locus also appears to be proportional to the product of the lengths of 5' and 3' flanking targeting homology.

This apparent dependence of targeted recombination frequency on the lengths of 5' and 3' flanking targeting homology (and not just the overall length of targeting homology) may reflect (i) the minimal unit length of uninterrupted homology required to effectively find its homologous genomic target sequence, form an initiation complex, and allow stable heteroduplex formation; (ii) a requirement for sufficient uninterrupted targeting sequence homology flanking each side of the

target gene deletion to allow initiation and extension of conversion tracts past the exon 5 *Mbo*II site or to facilitate resolution of crossovers on both arms of a replacement sequence flanking the deletion site (13, 15); and (iii) the necessity of maintaining these sufficient flanking homologies on each side of the target gene deletion despite rapid (and in some cases, quite extensive) intracellular exonucleolytic degradation of targeting fragment sequences (15, 16), and competition for these recombination substrates by both homologous and non-homologous extrachromosomal recombination pathways (26). The degree of involvement and relative impact of such competitive, extrachromosomal, degradative and recombinational processes might explain the longer homology lengths required for targeted recombination in mammalian cells compared with those required for intrachromosomal recombination between direct repeats in either mammalian cells or yeast (18, 19, 39).

While a number of previous studies have examined the homology dependence of targeted recombination in mammalian cells (9, 10, 14–17, 20, 30, 35, 36, 37), most of these studies have used selection or enrichment strategies that do not allow the recovery of certain recombinant classes (such as vector correction events), relatively few recombinants have been analyzed to determine the structure of the targeted recombination events, and in some cases, conflicting results have been reported by different laboratories. Several studies that have used linear sequence replacement vectors to target the *HPRT* locus in mouse ES cells (14–16) have reported that the majority of recombinants do not reflect simple gene replacement events but instead appear to involve integration of a recircularized or concatemericized vector into the target gene locus. Other studies (10, 36) have found that gene targeting with isogenic sequence replacement vectors appears to proceed very efficiently via simple gene replacement as long as sufficient uninterrupted flanking homology is provided on each arm of the targeting vector; however, when targeting homology on one arm of the vector is reduced to below 1 kb, many of the recombinants that are obtained reflect more complex and unorthodox vector concatemerization or vector integration events.

In this study, we have analyzed a total of 169 *APRT*⁺ recombinants, examining the effects of varying the length of targeting sequence homology on the types and relative frequencies of recombination events that were obtained using simple linear fragments of the Chinese hamster *APRT* gene as replacement-type targeting vectors. In experiments using *APRT* targeting fragments with at least 500 bp of homology flanking each side of the target gene deletion, we found that the majority of *APRT*⁺ recombinants (~53%) reflected simple targeted replacement or conversion events (Table 2). However, we observed a very marked and specific reduction in the relative frequency of these events when *APRT* fragments with less than 500 bp of flanking targeting homology on one side of the *Mbo*II site were used as targeting vectors.

Ten *APRT*⁺ recombinants showed targeted insertions of *APRT* targeting sequences at the chromosomal *APRT* locus. Even though they represent only a small fraction (5 to 7%) of the recombinants obtained in our targeting experiments, such recombinants provide considerable insight into the intracellular processes and recombinational pathways involved in gene targeting in mammalian cells. Each of these targeted insertion events appears to have arisen either by single-crossover, reciprocal recombination of an end-joined, recircularized *APRT* targeting fragment or concatemer of targeting fragments with the target gene locus or by nonconservative, targeted replacement of the chromosomal *APRT* targeting fragments. This also appears to be the case for targeted insertion events which have

been obtained in several studies using sequence replacement vectors to target the *HPRT* locus in mouse ES cells (14–16, 36). We found that in each of the 10 targeted insertion events obtained in our *APRT* targeting experiments, both terminal restriction sites have been lost from the inserted targeting fragment sequences, and that in the most of these recombinants, the inserted sequence appears to be ~50 to ~900 bp shorter than would be predicted based on the size of the *APRT* targeting fragment. These findings suggest that considerable exonucleolytic degradation of the ends of targeting fragment molecules must have taken place prior to any end joining or formation of the circularized or concatemericized recombinational intermediates that were ultimately integrated into the CHO-ATS49tg target gene locus. Such rapid and extensive intracellular exonucleolytic degradation of targeting vector sequences would effectively increase the apparent homology length requirements for efficient gene targeting.

Almost half of the *APRT*⁺ recombinants obtained in our *APRT* targeting experiments appear to have arisen by a novel recombinational process involving target gene-templated extension and correction of a 5'-truncated *APRT* targeting fragment sequence. Such targeting sequence correction recombinants have been previously obtained in *APRT* targeting experiments using both insertion-type and replacement-type targeting vectors, and with closed circular vectors as well as vectors which had been linearized either within or at the 5' or 3' end of homology (2, 3, 6, 23, 25). Similar events have also been detected in some ES cell targeting studies (32). Such events would not be predicted by conventional models of recombination (12, 33, 34), which typically require homology on both sides of a double-strand break.

One possible mechanism for such events would involve 3'-ended strand invasion by the *APRT* targeting sequence into the target gene duplex. This free 3' end could then prime DNA synthesis for its own extension, using the target gene sequence as a template (2, 3, 6). Similar strand invasion and invading strand-primed DNA synthesis events represent a major mechanism for late DNA replication in phage T4 (11, 22). The recombination structures generated by such one-sided invasion events could be resolved by unwinding of the newly synthesized strand or its displacement by branch migration, as well as by single-strand nicking. With the *APRT* fragments used as targeting vectors in this study, even relatively short (0.3- to 1.7-kb) extensions of the invading sequence would allow acquisition of the missing exon and promoter sequences necessary to generate a functional, WT copy of the *APRT* gene. However, our analyses indicate that these extensions are generally quite long; in most cases, the targeting sequences appear to have acquired an *Eco*RI site that is ~6 kb upstream of the *APRT* promoter. In several respects, the target gene-templated extension events that we see at the CHO *APRT* locus show a remarkable similarity to recombination events involved in the copying of sequences from ectopic sites via P-element-induced gap repair in *Drosophila melanogaster* (24).

Under conditions of limited 5' or 3' flanking homology, legitimate targeted gene replacement or conversion events constitute only a small fraction of the *APRT*⁺ recombinants obtained in CHO-ATS49tg cells; nearly 70% of the recombinants arise by target gene-templated extension and correction of targeting fragment sequences. The prevalence of targeting fragment correction under such circumstances suggests that templated extension events are much less dependent on the length of flanking homology than are gene conversion or replacement events, which appear to require substantial regions of homology on each side of the site being corrected. Targeting fragment correction events may require only enough

homology to initiate strand invasion and prime DNA synthesis; 5' flanking homology might not have to extend much beyond the exon 5 *Mbo*II site itself.

One could envision templated extension of the invading targeting fragment strand proceeding by a bubble migration mechanism similar to those described in models by Formosa and Alberts (11) and Nassif et al. (24). In such a model, templated extension of the invading strand would create only a relatively short D loop or moving replication bubble, with DNA synthesis driving strand displacement at the leading edge of the bubble and branch migration and reannealing of the target gene duplex at the trailing end rapidly displacing the newly synthesized strand and driving bubble migration. Templated extension of the invading strand would continue as long as the rate of DNA synthesis keeps pace with the rate of reannealing or until the DNA polymerase hits a pause site, allowing complete displacement of the extending strand by branch migration. Such a model would predict that target gene-templated extension and correction of targeting fragment sequences (i) should not require the formation of long, stable stretches of heteroduplex, (ii) should require only enough 3' flanking homology to initiate strand invasion and prime DNA synthesis and ensure inclusion of the *APRT* polyadenylation sequences, (iii) would require sufficient 5' flanking homology to protect the targeting fragment *Mbo*II site sequences from exonucleolytic degradation, and (iv) might result in the copying of very long stretches of upstream genomic sequences onto the ends of the invading sequence. For targeting fragments with limited flanking homology, such a mechanism would be expected to predominate over gene conversion or replacement events, since shorter 5' or 3' flanking homologies could decrease the likelihood of productive resolution of conversion tracts or the probability of forming heteroduplexes stable enough for crossovers to occur on both arms of a replacement sequence flanking the target gene deletion.

ACKNOWLEDGMENTS

We thank Richard Axel, Rodney Nairn, Susan Layher, Christi Walter, and John Wilson for kindly providing plasmids used to subclone or isolate the *APRT* targeting fragments for this study, and we thank Rodney Nairn, John Wilson, Debra Smith, and Robin Schneider-Broussard for helpful discussions and critical reading of the manuscript. We also thank Amir Salim and Marina Zabelshansky for technical assistance, Judy Ing and John Riley for assistance with photography and artwork, and Carol Hildman for help in manuscript preparation.

This work was supported by Public Health Service grant CA 28711 from the National Institutes of Health.

REFERENCES

- Adair, G. M., R. S. Nairn, J. H. Wilson, J. B. Scheerer, and K. A. Brotherman. 1990. Targeted gene replacement at the endogenous *APRT* locus in CHO cells. *Somatic Cell Mol. Genet.* **16**:437-441.
- Adair, G. M., R. S. Nairn, J. H. Wilson, M. M. Seidman, K. A. Brotherman, C. MacKinnon, and J. B. Scheerer. 1989. Targeted homologous recombination at the endogenous adenine phosphoribosyltransferase locus in Chinese hamster cells. *Proc. Natl. Acad. Sci. USA* **86**:4574-4578.
- Adair, G. M., J. B. Scheerer, D. G. Smith, K. A. Brotherman, S. Mason, M. M. Seidman, J. H. Wilson, and R. S. Nairn. Targeted recombination of replacement and insertion vectors at the endogenous *APRT* locus in Chinese hamster cells. Submitted for publication.
- Adair, G. M., R. L. Stallings, R. S. Nairn, and M. J. Siciliano. 1983. High-frequency structural gene deletion as the basis for functional hemizygosity of the adenine phosphoribosyltransferase locus in Chinese hamster ovary cells. *Proc. Natl. Acad. Sci. USA* **80**:5961-5964.
- Ahn, B. Y., K. J. Dornfield, T. J. Fagrellius, and D. M. Livingston. 1988. Effect of limited homology on gene conversion in a *Saccharomyces cerevisiae* plasmid recombination system. *Mol. Cell. Biol.* **8**:2442-2448.
- Aratani, Y., R. Okazaki, and H. Koyama. 1992. End extension repair of introduced targeting vectors mediated by homologous recombination in mammalian cells. *Nucleic Acids Res.* **20**:4795-4801.
- Ayares, D., L. Chekuri, K.-Y. Song, and R. Kucherlapati. 1986. Sequence homology requirements for intermolecular recombination in mammalian cells. *Proc. Natl. Acad. Sci. USA* **83**:5199-5203.
- Bradley, A. 1993. Gene targeting in embryonic stem cells. *Methods Enzymol.* **225**:855-879.
- Capecchi, M. R. 1989. Altering the genome by homologous recombination. *Science* **244**:1288-1292.
- Deng, C., and M. R. Capecchi. 1992. Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol. Cell. Biol.* **12**:3365-3371.
- Formosa, T., and B. M. Alberts. 1986. DNA synthesis dependent on genetic recombination: characterization of a reaction catalyzed by purified bacteriophage T4 proteins. *Cell* **47**:793-806.
- Haber, J. E., B. L. Ray, J. M. Kolb, and C. I. White. 1993. Rapid kinetics of mismatch repair of heteroduplex DNA that is formed during recombination in yeast. *Proc. Natl. Acad. Sci. USA* **90**:3363-3367.
- Hastings, P. J., C. McGill, B. Shafer, and J. N. Strathern. 1993. Ends-in vs. ends-out recombination in yeast. *Genetics* **135**:973-980.
- Hasty, P., J. Rivera-Pérez, and A. Bradley. 1991. The length of homology required for gene targeting in embryonic stem cells. *Mol. Cell. Biol.* **11**:5586-5591.
- Hasty, P., J. Rivera-Pérez, and A. Bradley. 1992. The role and fate of DNA ends for homologous recombination in embryonic stem cells. *Mol. Cell. Biol.* **12**:2454-2474.
- Hasty, P., J. Rivera-Pérez, C. Chang, and A. Bradley. 1991. Target frequency and integration pattern for insertion and replacement vectors in embryonic stem cells. *Mol. Cell. Biol.* **11**:4509-4517.
- Jiang, L., A. Conner, and M. J. Shulman. 1992. Effects of mutation position on frequency of marker rescue by homologous recombination. *Mol. Cell. Biol.* **12**:3609-3613.
- Jinks-Robertson, S., M. Michelitch, and S. Ramcharan. 1993. Substrate length requirements for efficient mitotic recombination in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **13**:3937-3950.
- 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.
- Miller, C. C. J., J. C. McPheat, and W. J. Potts. 1992. Targeted integration of the *Ren-1D* locus in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **89**:5020-5024.
- Miller, S. A., D. D. Dykes, and H. F. Polesky. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**:1215.
- Mosig, G. 1987. The essential role of recombination in phage T4 growth. *Annu. Rev. Genet.* **21**:347-371.
- Nairn, R. S., G. M. Adair, T. Porter, S. L. Pennington, D. G. Smith, J. H. Wilson, and M. M. Seidman. 1993. Targeting vector configuration and method of gene transfer influence targeted correction of the *APRT* gene in Chinese hamster ovary cells. *Somatic Cell Mol. Genet.* **19**:363-375.
- Nassif, N., J. Penney, S. Pal, W. R. Engels, and G. B. Gloor. 1994. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Mol. Cell. Biol.* **14**:1613-1625.
- Pennington, S. L., and J. H. Wilson. 1991. Gene targeting in Chinese hamster ovary cells is conservative. *Proc. Natl. Acad. Sci. USA* **88**:9498-9502.
- Roth, D., and J. Wilson. 1988. Illegitimate recombination in mammalian cells, p. 621-653. *In* R. Kucherlapati and G. R. Smith (ed.), *Genetic recombination*. American Society for Microbiology, Washington, D.C.
- Rothstein, R. 1991. Targeting, disruption, replacement and allele rescue: integrative DNA transformation in yeast. *Methods Enzy-*

- mol. **194**:281–301.
28. **Rubnitz, J., and S. Subramani.** 1984. The minimum amount of homology required for homologous recombination in mammalian cells. *Mol. Cell. Biol.* **4**:2253–2258.
29. **Shen, P., and H. V. Huang.** 1986. Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. *Genetics* **112**:441–457.
30. **Shulman, M. J., L. Nissen, and C. Collins.** 1990. Homologous recombination in hybridoma cells: dependence on time and fragment length. *Mol. Cell. Biol.* **10**:466–4472.
31. **Singer, B. S., L. Gold, P. Gauss, and D. H. Doherty.** 1982. Determination of the amount of homology required for recombination in bacteriophage. *Cell* **31**:25–33.
- 31a. **Smith, D. G., and G. M. Adair.** Unpublished data.
32. **Soriano, P., C. Montgomery, R. Geske, and A. Bradley.** 1991. Targeted disruption of the *c-src* proto-oncogene leads to osteoporosis in mice. *Cell* **64**:693–702.
33. **Sugawara, N., and J. E. Haber.** 1992. Characterization of double-strand break-induced recombination: homology requirements and single-stranded DNA formation. *Mol. Cell. Biol.* **12**:563–575.
34. **Szostak, J. W., T. L. Orr-Weaver, R. J. Rothstein, and F. W. Stahl.** 1983. The double-strand-break repair model for recombination. *Cell* **33**:5–35.
35. **te Riele, H., E. R. Maandag, and A. Berns.** 1992. Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proc. Natl. Acad. Sci. USA* **89**:5128–5132.
36. **Thomas, K. R., C. Deng, and M. R. Capecchi.** 1992. High-fidelity gene targeting in embryonic stem cells by using sequence replacement vectors. *Mol. Cell. Biol.* **12**:2919–2923.
37. **van Deursen, J., and B. Wieringa.** 1992. Targeting of the creatine kinase M gene in embryonic stem cells using isogenic and nonisogenic vectors. *Nucleic Acids Res.* **20**:3815–3820.
38. **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.
39. **Waldman, A. S., and R. M. Liskay.** 1988. Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. *Mol. Cell. Biol.* **8**:5350–5357.
40. **Wang, Q., and M. W. Taylor.** 1993. Correction of a deletion mutant by gene targeting with an adenovirus vector. *Mol. Cell. Biol.* **13**:918–927.
41. **Watt, V. M., C. J. Ingles, M. S. Urdes, and W. J. Rutter.** 1985. Homology requirements for recombination in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **82**:4768–4772.
42. **Yuan, L.-W., and R. L. Keil.** 1990. Distance-independence of mitotic intrachromosomal recombination in *Saccharomyces cerevisiae*. *Genetics* **124**:263–273.