End extension repair of introduced targeting vectors mediated by homologous recombination in mammalian cells

Yasuaki Aratani, Risa Okazaki and Hideki Koyama*

Kihara Institute for Biological Research, Yokohama City University, Nakamura-cho 2-120-3, Minami-ku, Yokohama 232, Japan

Received June 26, 1992; Revised and Accepted August 19, 1992

ABSTRACT

We have studied the mechanism of targeted recombination in mammalian cells using a hemizygous adenine phosphoribosyltransferase-deficient (APRT⁻) Chinese hamster ovary (CHO) cell mutant as a recipient. Three structurally different targeting vectors with a 5' or a 3', or both, end-deleted *aprt* sequence, in either a closed-circular or linear form, were transfected to the cells with a mutated *aprt* gene by electroporation. APRT-positive (APRT⁺) recombinant clones were selected and analyzed to study the gene correction events of the deletion mutation. Some half of 58 recombinant clones obtained resulted from corrections of the deleted chromosomal *aprt* gene by either gene replacement or gene insertion, a mechanism which is currently accepted for homologous recombination in mammalian cells. However, the chromosomal sequence in the remaining half of the recombinants remained uncorrected but their truncated end of the *aprt* gene in the incoming vectors was corrected by extending the end beyond the region of homology to the target locus; the corrected vector was then randomly integrated into the genome. This extension, termed end extension repair, was observed with all three vectors used and was as far as 4.6-kilobase (kb) or more long. It is evident that the novel repair reaction mediated by homologous recombination, in addition to gene replacement and gene insertion, is also involved in gene correction events in mammalian cells. We discuss the model which may account for this phenomenon.

INTRODUCTION

Gene targeting by means of homologous recombination between endogenous chromosomal locus and exogenous DNA is useful to modify the chromosomal gene into a designed structure (1). This contributes to investigations on the mechanism of gene expression, developments of animal models for human genetic diseases, and tests for the possibility of gene therapy.

Studies on the mechanism of targeted recombination in cultured mammalian cells may help to develop more improved technology for precisely and efficiently manipulating the mammalian genome and also provide insights into mechanisms of genetic rearrangements. In previous reports, recombination mechanisms have been investigated by exploring the recombination events between artificial (2, 3) or normal chromosomal target genes (4–8) and targeting vectors. These studies have demonstrated that most recombinants are generated by gene replacement and/or gene insertion. These mechanisms are well explained by the models proposed from studies on genetic recombination in microbial cells (9, 10).

We have investigated gene correction events at the deleted *aprt* locus in a CHO cell line. This is one of the ideal systems because APRT⁺ recombinants are directly obtained with a selective adenine/azaserine-containing medium. Three targeting vectors with either a 5' or a 3', or both, end-deleted *aprt* sequence were transfected and the resulting APRT⁺ recombinants were analyzed. We observed an unexpected recombination event in which the end-deleted targeting vectors were corrected by extending their ends beyond the region of homology between target locus and targeting vector. A similar phenomenon has been reported by Adair *et al* (8). We expand their observations and discuss the mechanism of the end extension reaction.

MATERIALS AND METHODS

Vector construction

Plasmid pD422 carrying a 3.9-kb *Bam*HI insert of the wild-type *aprt* gene cloned from the D422 cell line (fig. 1A) was a gift of M. Meuth (11). Targeting vector pAY1 (fig. 1B) was constructed from pD422 by cutting out a 0.6-kb fragment between the *Sma*I site in the multicloning site of plasmid pUC8 and the unique *Eco*RV site in exon 2 of the *aprt* gene and by ligating with an *Eco*RV linker. pOR1 (fig. 1B) was constructed from pD422 by removing a 2.0-kb sequence from the *Pst*I site in intron 3 to the *Pst*I site in the multicloning region flanking the 3' end of the *aprt* gene. pAE1 (fig. 1B) was made from pD422 by isolating a 1977-base pair (bp) *Pvu*II fragment containing exons 1 to 5 of the *aprt* gene and by ligating to the *Pvu*II site of pUC8.

* To whom correspondence should be addressed

Cells and culture conditions

CHO cell lines used in this study were provided by M. Meuth (11). D422 is the wild-type cell line which is hemizygous for the *aprt* gene and proficient in APRT activity. S118 is a spontaneous APRT⁻ mutant line having a 422-bp deletion ranging from intron 2 to exon 3 in the *aprt* locus (11) (indicated by open arrowheads in fig. 1A). Cells were routinely maintained as monolayers in 100-mm plastic dishes containing ES medium (Nissui Seiyaku Co.) (12) supplemented with 5% fetal bovine serum (Hyclone)(growth medium). All cultures were incubated at 37°C in humidified-air containing 5% CO₂. For the selection of APRT⁺ colonies, adenine (10⁻⁴M) and azaserine (2×10⁻⁵M) were added to ES medium supplemented with 5% dialyzed fetal bovine serum (AA medium) (12). In this selective medium, APRT⁺ cells are able to grow (AA-resistant; AA^r) but APRT⁻ are not (AA-sensitive; AA^s).

DNA transfection

DNA transfection was carried out by electroporation. Recipient S118 cells were grown to semi-confluence, harvested by trypsinization, washed once with Saline G buffer (13) and suspended at about 10⁸ cells/ml in the same buffer. A 40 µl aliquot of the cell suspension was exposed to 4 pulses at 3 KV/cm in the presence of targeting plasmid (4 µg) in a Shimadzu GTE-1 apparatus (Shimadzu Co.). After 15 min of incubation, the cells were diluted with growth medium, inoculated into 100-mm dishes each containing 10 ml of growth medium at 1–1.5×10⁵ cells/dish and cultured in a CO₂ incubator. Forty-eight hr later, the growth medium was changed to AA medium, followed by incubation for 14 days. Resulting colonies were picked, transferred to fresh AA medium and grown to mass cultures. The frequency of the AA^r (APRT⁺) colonies was calculated as a function of surviving cells, which was estimated by incubating aliquots of the electroporated cells in growth medium for 7 days. As controls, S118 cells (over 10⁸) were mock-transfected without added targeting DNA under the same conditions but no such colonies arose (data not shown). On the other hand, all targeting vectors failed to transform APRT⁻ S10 cells lacking the whole *aprt* sequence (11) into the enzyme-positive cells (data not shown).

Southern blot analysis

Total genomic DNA was extracted by the method of proteinase K digestion and phenol extraction in the presence of sodium dodecyl sulfate (SDS) (14). These DNA samples were digested to completion with restriction enzymes under reaction conditions recommended by the manufacturer (Takara Co.), electrophoresed in 0.8% agarose gels in 40 mM Tris-acetate buffer (pH 8.0), and then transferred to nylon membrane (Hybond N⁺, Amersham Japan). The 3.9-kb *Bam*HI insert containing the full length, wild-type *aprt* gene and a 0.6-kb *Bam*HI/*Eco*RV fragment which is not present in pAY1 were produced from pD422, labeled with [α-³²P]dCTP (~3000 Ci/mmol, Amersham Japan) by a random-primed DNA labeling kit (Boehringer Mannheim) to a specific activity of more than 10⁸ cpm/µg DNA and used for hybridization as an *aprt*-specific probe. Prehybridization was performed for 4 hr at 65°C in 6×SSC [0.9 M NaCl and 90 mM sodium citrate (pH 7.0)] containing 5×Denhardt's reagent (0.1% each of ficoll, polyvinylpyrrolidone and bovine serum albumin), and 100 µg denatured herring sperm DNA per ml. Hybridization was then performed for 18 hr at 65°C in the same buffer containing 10⁷ cpm of the ³²P-labeled probe per ml. The blots

were washed once with 1×SSC/0.5% SDS warmed at 65°C for 1 hr, followed by 30 min wash with 0.1×SSC/0.5% SDS in a 65°C water bath. Autoradiography was carried out with X-ray films (Kodak) at -80°C for 2–5 days.

PCR analysis

Genomic DNA (50 ng) was added to a reaction mixture (30 µl) containing 6 µl of 10×reaction buffer [500 mM KCl, 100 mM Tris-HCl (pH 9.0), and 15 mM MgCl₂], 0.1 mM each of 4 dNTPs, 30 ng each of two oligonucleotide primers, and 0.75 units of Taq DNA polymerase (5000 units/ml, Promega Co.). Primers for the reaction were derived from the *aprt* sequence. Primer I (5'-CAGGGGCTGCACCAAAGTGT-3') and II (5'-TCGCTTCCGGATGAGCACAC-3') were made to sequences flanking the region deleted in S118 cells. Primer III (5'-TCTCCTCGTGCTGGATCGCT-3') was made to the sequence 120-bp upstream of the transcription initiation site of the *aprt* gene absent in pAY1, while primer IV (5'-GCCAGGAGGAACCACTAGGA-3') was made to the sequence which is deleted in S118 cells. Samples were overlaid with 30 µl of mineral oil (Sigma); PCR was performed for 29 cycles of denaturation (94°C, 1 min), annealing (60°C, 2 min), and elongation (74°C, 3 min). Ten microliter aliquots of the products were resolved on a 2% agarose gel, stained with ethidium bromide (1 µg/ml) and amplified, diagnostic bands were identified.

RESULTS

Targeted correction of the deleted *aprt* gene

Recipient S118 cells are hemizygous for the *aprt* gene and totally deficient in APRT activity (AA^s), because the cells have a 422-bp deletion ranging from the intron 2 to the exon 3 of the *aprt* gene (fig. 1A) (11). Targeting vectors used first were pAY1 in a closed-circular form, or linearized with either *Eco*RV or *Bam*HI (fig. 2A). When linearized, these vectors were cut at the junctions of pUC8 and *aprt* sequences. Each of these DNAs was electroporated to S118 cells and AA^r (APRT⁺) colonies were selected in AA medium, as described in MATERIALS AND METHODS.

To know the nature of the targeted recombination events, we analyzed the structure of the *aprt* genes in 41 independent recombinant clones by Southern hybridization using the 3.9-kb *Bam*HI fragment of the *aprt* gene as a probe. Results with the representative 11 clones are shown in fig. 2B. In the D422 line, *Bam*HI digestion gave a 3.9-kb band (lane 1) which contained the wild-type *aprt* gene sequence as reported (11), while a deleted, 3.5-kb fragment was seen in the S118 cells (lane 2). Although different band patterns were observed in these recombinants (lanes 3–13), these were roughly classified into three groups. Recombinants 1, 6, and 10 (lanes 3, 8, and 12) showed only a 3.9-kb restriction fragment like the D422 line without any other band (group 1), indicating that the deletion of the *aprt* locus in S118 cells was corrected with the corresponding sequence of the introduced vector. As reported by others (3, 5, 6, 15), we thought that the target locus would be corrected by two types of recombination mechanisms: gene replacement by gene conversion or by double reciprocal cross-overs and gene insertion by single cross-overs. We digested the genomic DNA of each clone with *Bgl*II or *Kpn*I plus *Pvu*II (for the restriction enzyme sites, see fig. 1A) and hybridized it to the same full length probe. We obtained the predicted restriction fragments (data not shown)

Figure 1. Location of a deletion mutation at the *aprt* locus in the APRT⁻ CHO cell line S118 and targeting vectors. A, D422 is the wild-type cell line carrying one copy of the *aprt* gene which resides in two *Bam*HI sites [3.9-kilobase (kb) long]. Mutant line S118 has a 422-base pair (bp) deletion from intron 2 to exon 3 in the *aprt* gene; Open arrowheads indicate the termini of the deleted region. B, Targeting vectors pAY1, pOR1, and pAE1 were constructed from pD422 carrying the 3.9-kb *Bam*HI insert containing the wild-type *aprt* gene as described in MATERIALS AND METHODS. pAY1 loses the sequence from the *Bam*HI site upstream of the promoter region to the unique *Eco*RV site in exon 2. pOR1 lacks the sequence downstream of the *Pst*I site in intron 3. pAE1 lacks both regions upstream of the *Pvu*II site in exon 1 and downstream of the first *Pvu*II site in exon 5. Stippled boxes indicate exons numbered according to the map of Nalbantoglu *et al.* (11), and thick open bars, introns. Thick heavy lines indicate the chromosomal sequence flanking the *aprt* gene. Broken line indicates the pUC8 plasmid sequence depicted linearly. Open arrows indicate the restriction sites used for linearization of vectors. Figures are not drawn to scale. H, *Hind*III; Bg, *Bgl*II; B, *Bam*HI; P, *Pvu*II; E, *Eco*RV; Ps, *Pst*I; K, *Kpn*I.

confirming that recombinants 1, 6, and 10 were generated by gene replacement. Recombinant 2 showed the 3.9-kb band with extra bands of 5.6 and 6.0 kb after *Bam*HI digestion (lane 4), we confirmed by the results of *Bgl*II or *Pvu*II plus *Kpn*I digestion that this recombinant had arisen by gene insertion (group 2)(data not shown). On the other hand, recombinants 3, 4, 5, 7, 8, 9, and 11 (fig.2B, lanes 5, 6, 7, 9, 10, 11 and 13) (group 3) were characteristic of carrying a common *Bam*HI band of 3.5 kb along with the 3.9-kb and/or extra bands with sizes over 3.9 kb, suggesting that the *aprt* locus deleted in the S118 cells was not corrected.

The genome of each clone was further analyzed using PCR (fig. 3). PCR primers, I and II, made to sequences flanking the region deleted in S118 cells were used. In D422 (lane 1) and S118 (lane 2) cells, a 596-bp or a 174-bp band was amplified, respectively. Recombinants in group 1 (lanes 3, 8, and 12) only amplified the 596-bp band, confirming that the 422-bp deletion in the recipient cells was corrected. In the group 2 recombinant (lane 4), both 174- and 596-bp bands were amplified, indicating that the deleted and wild-type sequences were present in the genome as predicted for a single cross-over event. All recombinants belonging to group 3 showed both amplified bands (lanes 5, 6, 7, 9, 10, 11, and 13). The presence of this shorter band indicates that the target locus was not corrected and the longer band appears to have been derived from the targeting vector itself integrated into the genome. These data are all consistent with the results by Southern blotting described above (fig. 2). Among the 41 recombinant clones analyzed, 13 fell into group 1, 3 into group 2, and the remaining 25 into group 3.

Figure 2. Southern blot analysis of genomic DNA from the recombinant clones obtained with pAY1. A, The clones were produced by transfection of S118 cells with a closed-circular (a), *Eco*RV-cut (b), or *Bam*HI-cut (c) pAY1. Symbols are the same as those used in fig. 1. B, Genomic DNA (10 µg) purified from individual recombinant clones was restricted with *Bam*HI. Southern hybridization was performed as described in MATERIALS AND METHODS, using the ³²P-labeled, 3.9-kb *Bam*HI probe (shown in panel A). As controls, *Bam*HI-cut genomic DNA of wild-type D422 (lane 1) and mutant S118 (lane 2) cells was also run. Recombinants 1 to 5 (lanes 3 to 7), 6 to 9 (lanes 8 to 11), or 10 and 11 (lanes 12 and 13) yielded by transfection with the closed-circular, *Eco*RV-cut, or *Bam*HI-cut pAY1, respectively.

Extension of the deleted sequence of pAY1

As the *aprt* locus of the group 3 recombinants remained uncorrected, they were expected to have restored APRT activity by correcting the deleted sequence of the targeting vector, which was probably integrated into random sites of the genome. To demonstrate this point directly, we analyzed the 5' region of the *aprt* gene in those clones by Southern blotting using a probe of the 0.6-kb *Bam*HI-*Eco*RV fragment which is absent in pAY1 (fig. 2A). The 3.9-kb and/or additional bands with different sizes over 3.9 kb were detected in the group 3 recombinants as shown in fig. 4A (lanes 5-7, 9-11, and 13). This pattern of hybridization is the same as those observed with the 3.9-kb *aprt* probe (fig. 2), demonstrating that the transfected pAY1 indeed regained its 5'-deleted sequence and integrated into random sites of the genome. Some recombinants exhibited extra bands instead of the 3.9-kb band (fig. 4A, lanes 6, 10, and 13). This may best be explained as follows. The 0.6-kb sequence deleted in pAY1 was not fully synthesized, or was partially digested by nucleases even if synthesized; as a result, the corrected vectors did not possess either or both of the *Bam*HI sites flanking the *aprt* gene before integration. Moreover, the fact that in recombinants 5 and 9 (lanes 7 and 11) both 3.9-kb and extra bands in addition to the 3.5-kb

Figure 3. PCR analysis of genomic DNA from the recombinant clones obtained with pAY1. Genomic DNA for Southern hybridization in fig. 2B was used for amplification by PCR as described in MATERIALS AND METHODS, and the products were electrophoresed in a 2% agarose gel. Two synthetic primers I (5'-C-AGGGGCTGCACCAAAGTGT-3') and II (5'-TCGCTTCCGGATGAGCAC-AC-3') were configured to amplify a 596-bp sequence of the *aprt* gene in wild-type D422 cells (lane 1) while a 174-bp sequence in S118 cells (lane 2). Symbols and the numbers at the top of the gel are the same as those in fig. 2.

were hybridized to the 0.6-kb probe indicates that two molecules of the introduced vector in a recipient cell had been corrected and integrated into two different sites of the chromosome. In recombinant 3 (lane 5), such multiple correction events of the vector might occur and the corrected vector molecules would be integrated into the genome without losing their *Bam*HI sites, so that the 3.9-kb band would be more intense than the 3.5-kb band.

PCR analysis was carried out for confirmation (fig. 4B). Primer III was made to the sequence 110-bp upstream of the transcription start site of the *aprt* gene that is lost in pAY1, while primer IV to the sequence that is deleted in recipient S118 cells. Amplification of the 1114-bp sequence between the two primers occurred with the D422 cell genome (lane 1), but neither with the recipient genome (lane 2) nor with the targeting vector (data not shown) as a template. All recombinants (lanes 3–9) revealed the same diagnostic band amplified, clearly demonstrating that the *aprt* sequence of the incoming vector in these cells extended at least up to the region corresponding to primer III. The *aprt* gene in CHO cells has GC-rich sequences containing promoter activity specific for house-keeping genes, and the region running from the transcription start site (+1) to –89 is sufficient for a full promoter activity (16). Therefore, the corrected plasmids could have enough APRT activity to survive in AA medium.

As pAY1 lacks the sequence from the promoter region to the beginning of exon 2 in the *aprt* gene so that its 5' end joins the pUC8 sequence nonhomologous to the chromosomal *aprt* sequence (fig. 1), we did not expect such clones in which the incoming plasmid sequence could be corrected by homologous recombination with the target gene. However, such recombinants

Figure 4. Southern hybridization and PCR analysis of the 5' truncated region of transfected pAY1. A, The same filter used in fig. 2B was rehybridized with a 0.6-kb *Bam*HI-*Eco*RV fragment of the *aprt* gene (fig. 2A). B, Recombinant clones in which the chromosomal sequence was thought not to have been corrected were analyzed by PCR. Two primers III and IV, 1114-bp apart, were used for amplification. pAY1 lacks the sequence corresponding to primer III (5'-TCTCCTCGTGGATCGCT-3'), and S118 cells lack the sequence corresponding to primer IV (5'-GCCAGGAGGAACCAGTAGGA-3'), as illustrated above the gel. PCR products were electrophoresed in a 2% agarose gel followed by staining with ethidium bromide. Lane 1, D422; lane 2, S118; lanes 3 to 9, recombinant 3, 4, 5, 7, 8, 9, and 11, respectively. Symbols are same as those in fig. 1.

accounted for more than half of the APRT⁺ clones analyzed. We call this recombination event 'end extension repair' and will discuss the possible mechanism below.

End extension repair towards the 3' end of the *aprt* gene

The genome of 7 recombinants produced by transfecting *Pst*I-linearized pOR1 was also analyzed by Southern blotting (fig. 5) as mentioned above. Two were classified into group 1 (lane 3), and the remaining 5 revealed the blot patterns of group 3 (lanes 4 and 5). These results were ascertained by PCR analysis (data not shown). Therefore, it should be noted that the end extension repair of the targeting vector also occurs towards the 3' end of the deleted *aprt* gene.

Among 10 recombinant clones obtained by transfecting a closed-circular pAE1, 5 gave a single 3.9-kb band (fig. 5, lane 8) and were classified into group 1 defined above. Other 3 gave the 3.9-kb band plus an extra band of 6.0 kb (lane 9) and revealed only a 596-bp but not 174-bp band by PCR analysis (data not

Figure 5. Southern blot analysis of genomic DNA from the recombinant clones obtained with pOR1 and pAE1. Genomic DNA (10 μ g) was restricted with *Bam*HI and was subjected to Southern hybridization as shown in fig. 2. Lanes 1 and 6, D422; lanes 2 and 7, S118. Recombinants 12 (lane 3), 13 (lane 4), and 14 (lane 5) were obtained with a *Pst*I-cut pOR1; on the other hand, recombinants 15 (lane 8), 16 (lane 9), and 17 (lane 10) with a closed-circular pAE1.

shown), indicating that they arose by gene replacement accompanied by random integration of the targeting vector. The remaining 2 gave a 3.5-kb and an extra band (lane 10) falling into group 3. PCR analysis revealed that the pAE1 vector indeed extended upstream of the *aprt* gene till or beyond the promoter region, but that, unexpectedly, the 3' truncated end of the vector remained unrepaired (data not shown). In other experiments, we transfected S10 cells (absent of both copies of the *aprt* gene) with a plasmid lacking the sequence downstream from the first *Pvu*II site residing in exon 5 of the wild-type *aprt* gene and obtained AA⁺ colonies at almost the same frequency as that found after transfection with the wild-type gene itself (data not shown). These results indicate that the sequence downstream of that *Pvu*II site is not essential for functional APRT activity.

Size of end extension repair

Figure 6 shows that the incoming vector extended far upstream of the *aprt* gene. By transfecting S118 cells with a closed-circular pAY1, we obtained 5 recombinant clones in which their genomic DNA restricted with *Bam*HI exhibited the 3.5- and 3.9-kb bands by Southern blotting (recombinant 3 is represented in lane 5 of fig. 2B). Southern blot analysis was carried out after digestion of the same DNAs with *Hind*III. This digestion gave a 7.9-kb band in D422 cells (fig. 6, lane 1) while a 7.5-kb band in recipient S118 cells (lane 2). Since pAY1 has the *Hind*III site, on the pUC8 sequence, locating close to the 3' *Bam*HI site of the *aprt* insert (see fig. 1), the 7.9-kb band should be observed when the 5' end of the targeting vector extends up to the *Hind*III site residing upstream of the *aprt* locus (see panel A in fig. 1). Both 7.5- and 7.9-kb bands were found in all 5 recombinant clones, two of which (recombinant 3 and 3-1) are shown in fig. 6 (lanes 4 and 5). The 7.5-kb band is diagnostic for the chromosomal *aprt* gene because its deletion mutation remained uncorrected as shown in figs 2 and 3. These results demonstrate that the transfected pAY1 extended its 5' end up to or beyond the *Hind*III site which locates 4.6 kb upstream of the *Eco*RV site in exon 2 of the *aprt* locus. We conclude that the end extension is an unexpectedly long reaction proceeding more than 4.6-kb. We also examined recombinant 4 described in fig. 2 and observed a 7.5-kb band plus an extra band of 15 kb (fig. 6, lane 6). This would be

Figure 6. Southern blot analysis of genomic DNA restricted with *Hind*III. Genomic DNA was digested with *Hind*III and hybridized with the 3.9-kb *aprt* probe (fig. 2A). Lane 1, D422; lane 2, S118; lane 3, a mixture of D422 and S118. Recombinants 3 (lane 4), 3-1 (lane 5) and 4 (lane 6) were obtained with a closed-circular pAY1.

explained by two ways: the end extension could not reach the *Hind*III site, or after the extension the vector would lose either or both of the restriction sites.

Frequency of gene correction

The overall frequency of AA⁺, APRT⁺ recombinant clones obtained and the fraction of corrected *aprt* locus and targeting vector are summarized in table 1. On the whole, the APRT⁺ clones occurred at frequencies ranging from 1.7 to 11×10^{-7} . These frequencies are almost the same as those reported by Adair *et al.* (8), who have studied targeted correction events using a similar *aprt* system. The variation was due to the structure of targeting vectors and was also affected by their cutting sites prior to transfection. The *Bam*HI-cut pAY1 revealed the highest frequency under the present experimental conditions. Interestingly, although the frequencies at which the locus was corrected with the *Eco*RV-cut and *Bam*HI-cut pAY1 was 2.3- and 7.7-fold higher, respectively, than that with its uncut form, those of correction of the incoming vector was the same (3×10^{-7}). Therefore, linearization of the vector would stimulate the correction event of the target locus more greatly than that of the vector itself. With the closed-circular pAY1 and pAE1, the locus was corrected at similar frequencies, but the pAY1 itself was corrected 10-fold more efficiently than the pAE1. Although the pAY1 and pAE1 have 3277 and 1555 bp, respectively, of homology to the chromosomal sequence, that difference in vector correction seems not to be correlated with this extent of homology, because a similar observation was made between pAE1 and pOR1 which share the same homology of 1555 and 1548 bp, respectively. We do not know why the frequency of corrected pAE1 vector is significantly low compared to other vectors examined.

DISCUSSION

Using the *aprt* gene with a deleted mutation in the CHO cell line as a target locus (fig. 1A) and three targeting vectors characteristic of having differently truncated ends (fig. 1B), we studied gene correction events by homologous recombination. The frequency of recombination varied from 1.7 to 11×10^{-7}

Table 1. Frequency of gene correction of the deleted *aprt* locus and targeting vector.

Targeting vector*	AA ^r , APRT ⁺ clones ($\times 10^7$)	Corrected <i>aprt</i> gene ($\times 10^7$)	
		Locus	Vector
pAY1/uncut	4.0 (24)	1.0 (6)	3.0 (18)
pAY1/ <i>EcoRV</i>	5.4 (7)	2.3 (3)	3.1 (4)
pAY1/ <i>Bam</i> HI	11.0 (10)	7.7 (7)	3.3 (3)
pOR1/ <i>Pst</i> I	2.3 (7)	0.7 (2)	1.6 (5)
pAE1/uncut	1.7 (10)	1.4 (8)	0.3 (2)

* Each vector in an uncut (closed-circular) or linearized form with the restriction enzymes indicated was transfected to S118 cells, and AA^r, APRT⁺ recombinant clones were selected as described in MATERIALS AND METHODS. The frequency was calculated as a function of surviving cells. The number of the recombinant clones is shown in parentheses.

depending on the structure of targeting vectors used and their cut sites with restriction endonuclease for linearization prior to transfection (table 1). A significant stimulation of recombination by cutting outside the region homologous to the target sequence was observed; this has also been seen in recombination between two *aprt* sequence-containing plasmids (17), in which the recombination frequency increased 3-fold when one plasmid with a 5'-deleted *aprt* sequence was cleaved outside the region homologous to the other one.

Group 1 and 2 recombinants (figs. 2 and 3) had corrected their chromosomal sequences by targeted recombination events mediated by either gene replacement or gene insertion as described before (3, 5, 6, 15). Group 3 recombinants were found to result from correction of introduced vectors which were then randomly integrated into the genome (fig. 4). Although similar observations that the sequence of the incoming DNA is replaced by the chromosomal sequence have been reported (2, 3, 18), our present results were unexpected because pAY1, pOR1, or pAE1, lost entirely a 5', a 3' or both, ends of the *aprt* gene. Current recombination models (9, 10) require the presence of homology between two DNA sequences on both sides of a recombining site to initiate a recombination reaction. Since our vectors share a homologous sequence with only one side of the chromosomal *aprt* gene, these seemed not to be adequate substrates for correcting their own deleted region. However, it should be noted that more than half of the recombinant clones resulted from converting the truncated *aprt* gene of the vectors to a completely functional one. Similar recombination events have been reported in which the correction of an end-deleted *aprt* gene (8) and LINE-1 (19) was observed. Adair *et al.* prepared the targeting vector lacking the entire 5' flanking region plus first two exons of the *aprt* gene and used it for transfection after linearizing at the unique *Pst*I site (in intron 3) within the region homologous between the vector and the locus. This vector resembles our *EcoRV*-cut pAY1 in that both vectors possessed 5' ends of the *aprt* sequence generated by the double strand break. Here, this repair was also observed in the closed-circular and *Bam*HI-cut pAY1 having no such double strand break at the 5' end of the *aprt* gene at frequencies equal to that with the *EcoRV*-cut vector (table 1). These results indicate that prior to transfection, linearizing closed-circular vectors is not required, probably because they would rapidly be cut by nucleases within cells as will be discussed below. We also found that the extension proceeds towards the 3' end of the gene (fig. 5) and that the reaction extends over 4.6 kb or more (fig. 6).

Figure 7. End extension repair model by homologous recombination. Two possible mechanisms are shown, and the process of recombination and resulting products are depicted. Thin lines indicate the duplex DNA of a targeting vector such as *EcoRV*-linearized pAY1. Thick lines indicate the duplex DNA of the genome including the *aprt* locus. Broken lines represent a repair synthesis of DNA. Open arrows represent the sites cut with endonucleases. Arrowheads represent the 3' ends of the duplex.

We may explain this phenomenon by two possible mechanisms as illustrated in fig. 7. One mechanism (A) is that the strand ending 3' at the site of the double strand break in the vector invades the homologous region of the *aprt* locus followed by its repair synthesis from the end using the chromosomal, complementary strand as a template, forming one Holliday junction as shown in the double-strand break repair model (10). Here, we assume that, at the same time, the strand ending 5' at the site of the double strand break in the vector is repaired probably by a primase/polymerase alpha complex which is thought to be involved in the synthesis of a lagging strand in DNA replication fork (20). Resolution of the Holliday junction by an endonuclease following branch migration would generate an end-extended molecule of the targeting vector. The other mechanism (B) assumes an occasional DNA replication to make an eye form near the target locus (21). Both endonuclease and exonuclease cuts of this replicating DNA would give rise to a single-stranded tail, which could then pair with that generated from the incoming vector, as supposed by the single-strand annealing model (22). Following a gap repair of the pairing strand and a proper action of nucleases, an end-extended molecule including a full length *aprt* gene could be produced. According to the onionskin model of gene amplification (27), the target gene is amplified by multiple reinitiation of DNA replication within a single cell cycle. If a targeted recombination event occurs between one of such target genes and a vector molecule, resolution of the amplified region would result in one corrected *aprt* gene and the increased number of the deleted gene. This situation may interpret the great intensity of the 3.5-kb band as compared with that of the 3.9-kb band shown in recombinant 7 (figs. 2B and 4A, lane 9).

If either (or both) of these mechanisms works, a heterologous sequence connected to the truncated *aprt* end of a vector should interfere with its end-extending reaction. However, the repair frequency of the vector was unaffected even if the pUC8 sequence

linked to the truncated end (table 1). This suggests that the *aprt* sequence at or near the junction to the pUC8 is readily cleaved by an endonuclease after transfection so that a free 5' end of the gene is generated to initiate end extension repair. Bedale *et al.* have reported (23) that the RecA protein facilitates DNA strand breaks at homology/heterology junction and promotes DNA strand exchange. Similar enzyme(s) would act to cleave the *aprt*/pUC8 junction to initiate the end extension repair in CHO cells.

For identifying rare targeted recombinants, PCR screening has often been used (24–26). A pair of PCR primers is made to the sequence of a selective marker such as the *neo^r* gene inserted to a targeting vector and to the genomic sequence of the target locus which is not included in the vector. Using these primers one can detect a junction product with an expected length which is amplified only with targeted cells. However, Southern blot analysis (unpublished data) (26) has revealed that not all of such PCR-positive cells are true recombinants. This may be caused by an end-extending reaction of the targeting vector used, since this repair reaction proceeds for several kilobases as observed above (fig. 6). PCR/sib selection of recombinant cells in gene targeting experiments should be taken this point into consideration.

ACKNOWLEDGMENTS

We thank Miss H.Kojima for her technical assistance. This work was supported in part by grants from the Science and Technology Agency of Japan and from the Ministry of Education, Science, and Culture of Japan.

REFERENCES

1. Capecchi, M.R. (1989) *Science* **244**, 1288–1292.
2. Thomas, K.R., Folger, K.R. and Capecchi, M.R. (1986) *Cell* **44**, 419–428.
3. Song, K.Y., Schwartz, F., Maeda, N., Smithies, O. and Kucherlapati, R. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6820–6824.
4. Smithies, O., Gregg, R.G., Boggs, S.S., Koralewski, M.A. and Kucherlapati, R.S. (1985) *Nature* **317**, 230–234.
5. Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S. and Smithies, O. (1987) *Nature* **330**, 576–578.
6. Thomas, K.R. and Capecchi, M.R. (1987) *Cell* **51**, 503–512.
7. Valancius, V. and Smithies, O. (1991) *Mol. Cell. Biol.* **11**, 4389–4397.
8. Adair, G.M., Nairn, R.S., Wilson, J.H., Seidman, M.M., Brothman, K.A., MacKinnon, C. and Scheerer, J.B. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 4574–4578.
9. Meselson, M.S. and Radding, C.M. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 358–361.
10. Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J. and Stahl F.W. (1983) *Cell* **33**, 25–35.
11. Nalbantoglu, J., Hartley, D., Phear, G., Tear, G. and Meuth, M. (1986) *EMBO J.* **5**, 1199–1204.
12. Koyama, H. and Kodama, H. (1982) *Cancer Res.* **42**, 4210–4214.
13. Hama-Inaba, H., Takahashi, M., Kasai, M., Shiomi, T., Ito, A., Hanaoka, F. and Sato, K. (1987) *Cell Struct. Funct.* **12**, 173–180.
14. Ayusawa, D., Shimizu, K., Koyama, H., Takeishi, K. and Seno, T. (1983) *J. Biol. Chem.* **258**, 48–53.
15. Bollag, R.J., Waldman, A.S. and Liskay, R.M. (1989) *Annu. Rev. Genet.* **23**, 199–225.
16. Park, J. and Taylor, M.W. (1988) *Mol. Cell. Biol.* **8**, 2536–2544.
17. Wong, E.A. and Capecchi, M.R. (1986) *Somat. Cell Genet.* **12**, 63–72.
18. Orr-Weaver, T.L., Szostak, J.M. and Rothstein, R.J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6354–6358.
19. Belmaaza, A., Wallenburg, J.C., Brouillette, S., Gusew, N. and Chartrand, P. (1990) *Nucl. Acids Res.* **18**, 6385–6391.
20. Tsurimoto, T., Melendy, T. and Stillman, B. (1990) *Nature* **346**, 534–539.
21. Varshavsky, A. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 3673–3677.
22. Lin, F.L., Sperle, K. and Sternberg, N. (1984) *Mol. Cell. Biol.* **4**, 1020–1034.
23. Bedale, W.A., Inman, R.B. and Cox, M.M. (1991) *J. Biol. Chem.* **266**, 6499–6510.
24. Joyner, A.L., Skarnes, W.C. and Rossant, J. (1989) *Nature* **338**, 153–156.
25. Zimmer, A. and Gruss, P. (1989) *Nature* **338**, 150–153.
26. Soriano, P., Montgomery, C., Geske, R. and Bradley, A. (1991) *Cell* **64**, 693–702.
27. Stark, G.R. and Wahl, G.M. (1984) *Annu. Rev. Biochem.* **53**, 447–491.