Published in final edited form as:

DNA Repair (Amst). 2011 February 7; 10(2): 188–198. doi:10.1016/j.dnarep.2010.10.011.

Effects of Varying Gene Targeting Parameters on Processing of Recombination Intermediates by ERCC1-XPF

Jennifer J. Rahn^{1,3}, Brian Rowley¹, Megan P. Lowery¹, Luis Della Coletta¹, Tiffany Limanni¹, Rodney S. Nairn^{1,2,*}, and Gerald M. Adair¹

¹ Department of Molecular Carcinogenesis, The University of Texas M.D. Anderson Cancer Center, Science Park - Research Division, P.O. Box 389, 1808 Park Road 1C, Smithville, TX 78957, USA

² The University of Texas Graduate School of Biomedical Sciences at Houston, P.O. Box 20334, Houston, TX 77225, USA

Abstract

The ERCC1-XPF structure-specific endonuclease is necessary for correct processing of homologous recombination intermediates requiring the removal of end-blocking nonhomologies. We previously showed that targeting the endogenous CHO *APRT* locus with plasmids designed to generate such intermediates revealed defective recombination phenotypes in *ERCC1* deficient cells, including suppression of targeted insertion and vector correction recombinants and the generation of a novel class of aberrant recombinants through a deletogenic mechanism. In the present study, we examined some of the mechanistic features of ERCC1-XPF in processing recombination intermediates by varying gene targeting parameters. These included altering the distance between the double-strand break (DSB) in the targeting vector and the inactivating mutation in the *APRT* target gene, and changing the position of the target gene mutation relative to the DSB to result in target mutations that were either upstream or downstream from the DSB. Increasing the distance from the DSB in the targeting vector to the chromosomal target gene mutation resulted in an *ERCC1* dependent decrease in the efficiency of gene targeting from intermediates presenting lengthy end-blocking nonhomologies. This decrease was accompanied by a shift in the distribution of recombinant classes away from target gene conversions to targeted insertions in both wild-type and *ERCC1* deficient cells, and a dramatic increase in the proportion of aberrant recombinants in *ERCC1* deficient cells. Changing the position of the target gene mutation relative to the DSB in the plasmid also altered the distribution of targeted insertion subclasses recovered in wild-type cells, consistent with two-ended strand invasion followed by resolution into crossover-type products and vector integration. Our results confirm expectations from studies of Rad10-Rad1 in budding yeast that ERCC1-XPF activity affects conversion tract length, and provide evidence for the mechanism of generation of the novel, aberrant recombinant class first described in our previous study.

*Correspondence to: Rodney S. Nairn, Ph.D., 512/237-9464 (voice), 512/237-6533 (fax), RNairn@mdanderson.org.

³Present Address: Medical University of South Carolina, Department of Pharmaceutical and Biomedical Sciences, 280 Calhoun Street, Charleston, SC 29425, USA

The authors declare that there are no conflicts of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

ERCC1; gene targeting; aberrant recombination; *APRT*

1. Introduction

ERCC1-XPF is a heterodimeric, structure-specific endonuclease functioning in nucleotide excision repair (NER) [1–4], the processing of specific types of intermediates generated during strand-invasion and single-strand annealing homologous recombination [5–7] and telomere maintenance [8,9]. In *Saccharomyces cerevisiae*, *RAD10* and *RAD1* are the *ERCC1* and *XPF* homologues, respectively, and together constitute the corresponding heterodimeric endonuclease Rad10-Rad1 [10–12]. The structure-specific nicking activity of these endonucleases has a 5' polarity with respect to the NER bubble or other double-stranded to single-stranded transitional structures used as substrates; that is, endonucleolytic nicking occurs on the 5' side of the junction at the transition from double-stranded to single-stranded DNA [13–16]. Experiments in budding yeast and mammalian cells have demonstrated the requirement for Rad10-Rad1 and ERCC1-XPF in precisely removing lengthy nonhomologous DNA sequences that block 3'-OH homologous ends during plasmid-chromosome gene targeting [5,17–19].

In a previous study [5], we targeted the endogenous, hemizygous *APRT* locus in a CHO cell line (ATS49tg) and an *ERCC1* knock-out cell line derived directly from it (E1KO7-5) with plasmids designed to interrogate the role of ERCC1-XPF in processing strand invasion recombination intermediates; the structures of these two targeting vectors are shown in Figure 1A. Both cell lines contained the same *APRT* mutant locus harboring an inactivating 3-bp deletion in exon 5 abolishing a *Mbo*II restriction site. In these experiments examining targeted gene correction by plasmid-chromosome recombination, the mutation in the *APRT* target locus was positioned approximately 450-bp downstream from the double-strand break (DSB) in the targeting vector [5]. Loss of *ERCC1* resulted in dramatic reductions in the recovery of targeted insertion and vector correction recombinants, as well as the recovery of a significant number of aberrant recombinants, when the targeting vector (pAG6ins0.9) contained lengthy 3' end-blocking nonhomologous sequences. Such effects were not observed when a targeting vector with short end-blocking nonhomologous sequences (pAG6) was used. Results of this study implicated ERCC1-XPF directly in precisely removing lengthy nonhomologous DNA sequences flanking the DSB in the targeting vector, and established a requirement for ERCC1-XPF in the normal processing of such end-blocked recombination intermediates. Generalized structures of the types of recombinants expected and recovered from these experiments are represented in Figure 1B; Figure 2 (left side) shows the structure of the *APRT* target locus with the exon 5 *Mbo*II deletion used in this previous study [5].

Results from a number of investigations using *S. cerevisiae* have shown that the Rad10-Rad1 structure-specific endonuclease functions similarly to ERCC1-XPF in processing specific types of recombination intermediates in chromosome-plasmid gene targeting and other recombination pathways [17–20]. However, the generation of aberrant recombination events when the targeting vector with long terminal nonhomologies was used in the *ERCC1* null cell line, as documented in our previous experiments, has not been observed in yeast. In our previous study, recombinants having aberrant structures comprised 16% of the total of *APRT*⁺ recombinants. These aberrant recombinants resembled insertions of the targeting vector at the *APRT* locus, except that there were variable-sized deletions positioned downstream from the corrected *APRT* copy. We also recovered recombinants in which the targeting vector, containing either long or short stretches of nonhomologous sequence, was

repaired by recombination and then randomly integrated to confer the *APRT*⁺ phenotype; this recombination phenotype was also uniquely observed in our experiments in CHO cells [5]. This class of recombinant, termed vector correction, was recovered in wild-type cells using both targeting vectors. In *ERCC1* deficient cells, such vector correction recombinants were recovered only from the vector presenting short terminal nonhomologies (i.e., pAG6), indicating deficient gap repair in *ERCC1* deficient cells when 3'-OH homologous termini are blocked by long single-stranded heterologous tails after DNA end resection [5].

Recombination phenotypes associated with mutations in *RAD1* and *RAD10* include shortening of conversion tracts and reduction of conversion-associated crossovers, and deficient gene replacement and plasmid integration [21,22]. In the present study, we lengthened the distance between the DSB in the targeting vector and the small inactivating deletion in the *APRT* target gene from ~450-bp to ~1200-bp to determine if this increased distance would impact targeted recombination in our *ERCC1* deficient cell lines. At the same time, we altered this gene targeting system by using a different CHO cell line with the inactivating *APRT* mutation located upstream rather than downstream (as in our previous experiments) from the DSB in the targeting plasmid. By manipulating the location of the mutation in the targeted *APRT* locus in this way, we were able to confirm our previous speculation as to the mechanism of generating targeted insertion recombinants in this system, and also test the notion that the novel class of aberrant recombinants we reported in our previous study [5] arose from the abortive integration of the plasmid substrate. Using this approach, we show here that both wild-type and *ERCC1* deficient cells were less efficient at successfully completing gene targeting when the distance between the DSB in the plasmid and the target gene mutation was increased. In addition, several significant changes were observed in the distributions of recombinant classes. These differences depended on *ERCC1* functionality, targeting vector configuration, and the location of the inactivating mutation in the target locus, i.e., being either upstream or downstream from the DSB in the vector. (For conciseness, we will refer to this hereafter as the “targeting polarity.”) Reversing targeting polarity in this way caused a pronounced shift in the distribution of subclasses of targeted insertion recombinants in wild-type cells, and a significant increase in the generation of aberrant recombinants in *ERCC1* deficient cells targeted with the end-blocked plasmid. The structures of aberrant recombinants from the experiments reported here differed from those previously described [5] in that the deletions were in 5' rather than 3' orientation relative to the original location of the DSB in the targeting vector. Together these observations indicate that targeted insertions and aberrant recombinants are generated through a common mechanism, and that aberrant insertion recombinants likely arise through abortive plasmid integration in *ERCC1* deficient cells.

2. Materials and methods

2.1. Plasmids and cell lines

Plasmids pAG6 and pAG6ins0.9 were constructed and used in previous studies of targeted recombination in CHO cells [5,23]. These targeting vectors (shown in Figure 1A) each contain the entire CHO *APRT* gene on a 3.9 kb *Bam*HI fragment cloned into plasmid pSV2gpt [23]. In pAG6, there is a 33-bp polylinker insertion at the *Xho*I site in exon 3 of *APRT*, and for pAG6ins0.9 there is a 945-bp insertion (derived from a nonfunctional *Nae*I fragment of HSV-TK) at the same site; these insertions render the plasmid *APRT* sequences inactive. When digested with *Hind*III (for pAG6) or *Sal*I (for pAG6ins0.9), the plasmids are linearized and represent insertion-type (i.e., “ends-in”) targeting vectors in which 18- and 11-nucleotide (pAG6) or 271- and 670-nucleotide (pAG6ins0.9) nonhomologous sequences flank each side of the DSB after resection, blocking the ends of *APRT* sequences homologous to the target locus (see Figure 1A). The *GPT* (xanthine-guanine phosphoribosyltransferase) cassette in the plasmid backbone allows selection of HAT

(hypoxanthine-amethopterin-thymidine)-resistant colonies in hypoxanthine-guanine phosphoribosyltransferase (*HGPRT*)-deficient CHO cells [24] thereby allowing selection and scoring of non-targeted events (i.e., random integrations) in HAT medium.

CHO cell line ATS49tg is derived from the CHO AT3-2 cell line [25], and was selected stepwise for spontaneous resistance to 8-azaadenine (8-AA) and 6-thioguanine (6-TG), and thus contains spontaneous inactivation mutations in the *APRT* and *HGPRT* genes, respectively. The *APRT* mutation has been characterized as a 3-bp deletion in exon 5 abolishing a *Mbo*II restriction site [26]. ATS49tg is hemizygous for both the *APRT* and *ERCC1* loci [27,28], and cell line E1KO7-5 was derived directly from ATS49tg by targeted gene disruption of *ERCC1* using the positive-negative selection vector pSL1, as described [29]. These two cell lines were previously used in *APRT* targeted gene correction experiments [5] and generated a large number of *APRT*⁺ targeted correction recombinants, including gene conversion recombinants in which the exon 5 *APRT* mutant locus in ATS49tg and E1KO7-5 cell lines was converted to wild-type, restoring the *Mbo*II restriction site. To generate cell lines ATS49-Rv and E1KO-Rv, two such *APRT*⁺ convertants from each cell background were electroporated with restriction enzyme *Eco*RV as described [30] and 8-AA-resistant clones were screened by PCR for loss of the unique *Eco*RV site in *APRT* exon 2. Cell lines ATS49-Rv and E1KO-Rv were two clones each of which had an identical 2-bp deletion in exon 2 that inactivated *APRT* and resulted in the loss of a unique *Eco*RV site; this was verified by DNA sequencing of genomic DNA. ATS49-Rv and E1KO-Rv are thus nearly isogenic with each other and their respective parental cell lines ATS49tg and E1KO7-5; however, we can not rule out that additional mutations may have been introduced by the *Eco*RV electroporation step used in their generation. These cell lines were used in *APRT* targeted gene correction experiments described here with targeting vectors pAG6 and pAG6ins0.9 (see Figure 1).

2.2. Cell culture and selection conditions

Cell lines were cultured in alpha-modified Minimal Essential Medium (α -MEM) containing 2 mM L-glutamine, penicillin (50 U/ml), streptomycin (50 μ g/ml), gentamycin (50 μ g/ml) and 10% fetal calf serum (FCS) in a humidified atmosphere of 5%CO₂/95% air at 37 °C, unless otherwise stated. For selection of *APRT*⁺ clones after gene targeting, alanosine-azaserine-adenine (ALASA) selection medium was used, containing 25 μ M alanosine, 50 μ M azaserine, and 100 μ M adenine in α -MEM with 10% FCS. For selection of *GPT*⁺ clones, HAT media containing 100 μ M hypoxanthine, 2 μ M amethopterin, and 50 μ M thymidine in α -MEM with 10% FCS was used. L-alanosine (NSC-529469) was a gift of the Drug Synthesis and Chemistry Branch, National Institutes of Health.

2.3. Gene targeting

Procedures for gene targeting by electroporation have been described in detail [31]. Briefly, restriction endonuclease-digested vectors pAG6 and pAG6ins0.9 (using *Hind*III and *Sal*I, respectively) were treated with proteinase K (200 μ g/ml) in 0.1% SDS and 1 M lithium acetate, subjected to phenol extraction, re-isolated by precipitation from ethanol and suspended in sterile water. Subconfluent monolayers of cells were harvested after trypsin-EDTA treatment, washed, counted and re-suspended in ice-cold LEP electroporation buffer (272 mM sucrose, 1 mM MgCl₂, 7 mM sodium phosphate, pH 7.4) at 2×10^7 cells/ml. Using a Bio-Rad Gene Pulser Electroporator set at 25 μ F, 450 V, 0.85 ml cells (17×10^6) in Bio-Rad electroporation cuvettes (4-mm gap) were electroporated with 11 μ g/ml digested targeting vector and then diluted in 8.5 ml ice-cold serum-free α -MEM prior to plating into 100-mm dishes containing warm α -MEM with 10% FCS. After incubation at 37 °C for 40–48 hours to allow expression, plates were refed with either HAT or ALASA selection media, incubated for 8 – 10 days longer at 37 °C and enumerated. Individual colonies were picked,

expanded, and DNA was isolated; analysis of recombinant clone DNA structures by Southern blotting was performed as previously described [5,31].

3. Results

3.1. Frequencies of targeted recombination

Table 1 shows the frequencies of targeted correction of the endogenous *APRT* locus by homologous recombination in the four CHO cell lines targeted with vectors pAG6 and pAG6ins0.9; data for cell lines ATS49tg and E1KO7-5, targeting the exon 5 *APRT* mutant allele (i.e., *MboII*Δ), are from a previously published study [5], and are shown for comparison. These results indicate that the efficiency of targeting at the hemizygous CHO *APRT* locus (as determined by the ratio of targeted *APRT*⁺ recombinants to untargeted *GPT*⁺ random integrations) is lower for both ATS49-Rv and E1KO-Rv cell lines (bearing the *EcoRV* exon 2 mutation, *EcoRV*Δ) when compared to ATS49tg and E1KO7-5 cell lines harboring the *MboII*Δ allele. The only differences between these two sets of cell lines are that (i) the polarity of targeting is reversed (from the inactivating mutation being located downstream from the DSB in the vectors in previous experiments to upstream in the present study), and (ii) the distance of the deletion in the *APRT* target is increased from ~ 450 to ~ 1200 bp. This reduction in targeting efficiency was observed regardless of the targeting vector used, although the targeting efficiency was reduced to a greater extent with the pAG6ins0.9 vector. The requirement of a longer conversion tract for productive *APRT* targeting in cell lines ATS49-Rv and E1KO-Rv (compared to ATS49tg and E1KO7-5) due to the greater distance of the target locus mutation from the DSB may explain this loss of efficiency. However, the fact that targeting efficiency is further reduced in pAG6ins0.9 compared to pAG6 targeting can not be explained by changes in conversion tract or targeting polarity. Since the wild-type cell line ATS49-Rv and *ERCC1* deficient cell line E1KO-Rv were not differentially affected in either pAG6 or pAG6ins0.9 targeting experiments, this observed reduction in targeting efficiency is apparently independent of *ERCC1*-XPF activity. These observations led to questions of whether or not *ERCC1* deficiency could affect the distribution of individual recombinant classes recovered from targeting experiments in which the distance from the DSB in the vector and the target mutation was increased, and the targeting polarity was reversed by positioning the mutation upstream rather than downstream from the DSB.

3.2 Structures of *APRT*⁺ recombinants

Figure 1 illustrates the structures of targeting vectors pAG6 and pAG6ins0.9 (Fig. 1A) and of the various classes of *APRT*⁺ targeted recombinants expected to be generated from targeted gene correction experiments in the cell lines used in this study: target gene conversions, targeted insertions, and targeting vector corrections (Fig. 1B). Additionally, Figure 2 shows the structures of subclasses of targeted insertion recombinants possible depending on the specific *APRT* locus targeted, i.e. *MboII*Δ (mutation downstream; left side of diagram; used in our previous study [5]) vs. *EcoRV*Δ (mutation upstream; right side of diagram; used in this study) targets. These include targeted insertions in which the upstream (5') duplicated *APRT* copy is wild-type and the downstream (3') duplicated *APRT* copy contains the original target deletion (+/-), insertions in which both *APRT* copies are wild-type (+/+), and insertions in which the upstream copy contains the original target mutation and the downstream copy is wild-type (-/+). In addition, recombinants with aberrant structures can be recovered which contained variable-sized deletions, as shown at the bottom of Figure 2 (+/Δ and Δ/+). In our previous study [5], these deletions extended only in the 3' direction downstream from a wild-type 5' *APRT* copy (i.e., +/Δ), and in this study, we recovered aberrant recombinants with deletions extending upstream in the 5' direction relative to a wild-type 3' *APRT* copy (i.e., Δ/+).

Figure 2 also serves to illustrate our experimental design for addressing the role of ERCC1-XPF in generating these various recombinant structures. As shown in the center of the diagram, the pathway leading to gene conversions resulting from processing of recombination intermediates arising between the targeting vector presenting lengthy end-blocking nonhomologies and the targeted *APRT* locus (either *MboII*Δ or *EcoRV*Δ) does not absolutely require ERCC1-XPF to generate gene conversion recombinants [5], although recent results [33,34] have suggested that ERCC1-XPF does affect conversion pathways in intrachromosomal recombination assays. However, the targeted integration pathway critically requires ERCC1-XPF to produce the heteroallelic duplicate targeted insertion recombinant structures shown on the left side of Figure 2, as demonstrated in our previous study [5] targeting the exon 5 *MboII*Δ *APRT* allele. In the absence of ERCC1-XPF activity, these recombinants were ablated and the aberrant structures shown at the bottom left were produced (+/Δ). As noted above, only +/+ and +/- targeted insertions and +/-Δ aberrant recombinants were obtained from these experiments. By reversing the “targeting polarity” through use of the *EcoRV*Δ *APRT* allele as a target (right side of diagram), in the present study we were able to address the mechanism for generating these structures, and test directly whether aberrant recombinants, occurring in the absence of ERCC1-XPF, result by faulty processing of recombination intermediates through the pathway normally leading to targeted insertions. The predictions and interpretation of the results from this experimental design are detailed in the “Discussion.”

As shown in Table 2, all the recombinant structures illustrated in Figures 1 and 2 were generated in the combined results of our previous study [5] and the study described here. The *APRT* locus in the CHO cell lines used in our targeting experiments is hemizygous, and therefore recombinant structures may be straightforwardly determined by analysis of restriction endonuclease fragments after Southern blotting. Figure 3 illustrates the various structures possible from targeting the exon 2 *EcoRV*Δ *APRT* allele in cell lines ATS49-Rv and E1KO-Rv used in the present study, and restriction digestion fragments generated from each used to determine their structures. Supplemental Figures 1–3 show examples of Southern blots of different *APRT* recombinants in cell lines ATS49-Rv and E1KO-Rv generated from targeting vectors pAG6 and pAG6ins0.9. Supplemental Figure 1 shows examples of pAG6 targeted recombinants in cell lines ATS49-Rv (panels A and B) and E1KO-Rv (panels C and D); Supplemental Figure 2 shows examples of pAG6ins0.9 targeted recombinants in cell lines ATS49-Rv (panels A and B) and E1KO-Rv (panels C and D); Supplemental Figure 3 shows examples of pAG6ins0.9-generated aberrant recombinants in cell line E1KO-Rv. As in our previous study targeting the exon 5 *MboII*Δ allele in cell lines ATS49tg and E1KO7-5 [5], only *ERCC1* deficient cells targeted with pAG6ins0.9 generated aberrant recombinants. In all the aberrant recombinants recovered from E1KO-Rv, variable-sized deletions extended upstream from a 3' wild-type copy of *APRT*, opposite in orientation to those recovered from E1KO7-5 [5]. The deduced structures of representative aberrant recombinants from this study (corresponding to those clones analyzed in Supplemental Figure 3) are shown in Figure 4.

3.3. Class distributions of *APRT*⁺ recombinants

All classes of the recombinants shown in Figures 1, 2 and 3 were recovered in the experiments reported here and previously [5]. Notably, a number of changes were apparent in the class distributions of recombinants when comparing targeting results between the different cell lines from the two different targeting vectors, as shown in Table 2. Table 2 includes previously published data for cell lines ATS49tg and E1KO7-5 [5] to facilitate these comparisons. It should be noted that the data are constituted from Southern analyses of a large number of clones (e.g., up to 139 for the E1KO-Rv cell line targeted with pAG6ins0.9) isolated after ALASA selection from the different experimental conditions,

and that the experimental design ensured that each clone represented an independent recombination event [5, 31, 32], facilitating statistical comparisons. All p values were calculated by one-sided Fisher's exact tests. Specific differences in the recombinant class distributions shown in Table 2 and the interpretation of these differences in context of their mechanistic implications are discussed below.

4. Discussion

This study was designed to address some aspects of the mechanisms of gene targeting in mammalian cells, and to extend our understanding of the role of the structure-specific endonuclease ERCC1-XPF in this process. Results of a previous study [5] demonstrated that productive gene targeting required ERCC1-XPF to precisely remove lengthy nonhomologous DNA sequences blocking the ends of homology in targeting vector DNA. *ERCC1* deficient cells demonstrated several recombination defective phenotypes when targeted with such a vector, including dramatically reduced frequencies of targeted insertion and target vector correction recombinants, and the generation of novel, aberrant recombinants, not observed in isogenic *ERCC1* proficient cells, which resembled targeted insertions but contained large, variable-sized deletions. These results indicated that ERCC1-XPF plays a critical role in resolving recombination intermediates in pathways leading to crossover-type products during gene targeting.

Results from recent studies of DSB-induced mitotic (i.e., intrachromosomal) homologous recombination in *ERCC1* mutant UV4 [33] and *XPF* mutant UV41 [34], verified the requirement for ERCC1-XPF in resolving recombination intermediates into crossover-type products, and in addition provided evidence that ERCC1-XPF was also required for efficient intrachromosomal gene conversion. In gene targeting, the pathway leading to gene conversions is likely also affected by ERCC1-XPF, since *ERCC1* deficient mouse embryonic stem cells are deficient in targeted gene replacement [35], reflecting inability to convert the target allele when “ends-out” as opposed to “ends-in” (as in this study) targeting vectors are used. In *Saccharomyces cerevisiae*, *rad1* mutants exhibit reduced conversion tracts and *RAD1*, the *XPF* homologue, is necessary for crossovers associated with gene conversion, which require long conversion tracts in yeast [21]. Based on these observations, we designed gene targeting experiments to address the effects of *ERCC1* deficiency on the mechanisms of resolving plasmid-chromosome recombination intermediates in mammalian cells.

We previously targeted an exon 5 mutant *APRT* allele (*MboII* Δ) in isogenic *ERCC1* proficient and deficient CHO cell lines, using “ends-in” configured targeting vectors presenting either very short (pAG6) or lengthy (pAG6ins0.9) end-blocking nonhomologies [5]. For the experiments reported here, we targeted an exon 2 mutant *APRT* allele (*EcoRV* Δ) located ~1200-bp (as opposed to ~450-bp for the *MboII* Δ allele) from the DSB in the targeting vectors, thus increasing the distance by about 2.7-fold for productive conversion if recombination is initiated from the DSB in the plasmid. Increasing this distance reduced the proportions of target gene conversions significantly in both wild-type and *ERCC1* deficient cell lines, independently of which targeting vector was used. As shown in Table 2, wild-type ATS49-Rv (*EcoRV* Δ) cells compared to wild-type ATS49tg (*MboII* Δ) cells showed decreases from 35% to 16% ($p=0.04$) and from 46% to 26% ($p=0.003$) when targeted with pAG6 and pAG6ins0.6, respectively. Also, *ERCC1* deficient E1KO-Rv (*EcoRV* Δ) compared to E1KO7-5 (*MboII* Δ) cells showed decreases from 41% to 17% ($p=0.03$) with pAG6 and from 82% to 38% ($p=0.001$) with pAG6ins0.9. These results indicate that, as expected, increasing the required distance for generating target gene conversions decreases their successful production, and that this does not depend on targeting vector configuration or *ERCC1*. Interestingly, however, although targeting either pair of isogenic cell lines with

plasmid pAG6 revealed no significant differences that depended on *ERCC1* functionality (i.e. 35% vs. 41% and 16% vs. 17%), targeting *ERCC1* deficient cells with pAG6ins0.9 produced significant increases in target gene conversions for both *APRT* alleles (to 82% from 46% for the *MboII*Δ allele, $p = 0.001$, and to 38% from 26% for the *EcoRV*Δ allele, $p = 0.03$). These increases can be partly explained by the fact that *ERCC1* deficient cells are defective in resolving recombination intermediates from pAG6ins0.9 into crossover-type recombinants, as discussed in more detail below, and that therefore more of the intermediates generated during homologous recombination may be channeled into gene conversions in *ERCC1* deficient cells. The *MboII*Δ *APRT* target allele in E1KO7-5 requires a shorter conversion tract (as determined by the distance required from the DSB break in the vector to the target gene mutation) to generate target gene conversions than does the *EcoRV*Δ *APRT* allele in E1KO-Rv, and the observation that this increase in gene conversions is less pronounced in E1KO-Rv may also reflect an effect of *ERCC1* deficiency itself in shortening productive conversion tracts, similar to the defect in *rad1* mutants [21,36].

Another effect of increasing the distance from the vector DSB to the target mutation can be observed by considering the different proportions of targeting vector correction recombinants generated in cell lines with different target alleles (i.e. *MboII*Δ in ATS49tg/E1KO7-5 and *EcoRV*Δ in ATS49-Rv/E1KO-Rv). These recombinants represent events in which the nonhomologous sequences flanking the DSB in the plasmid are removed and the gap is precisely repaired by recombination using sequence information from exon 3 of the *APRT* target allele (see Fig. 1 and Fig. 3E). As shown in Table 2, for pAG6 targeting there were no differences between wild-type and *ERCC1* deficient cells for either *APRT* allele (35% vs. 28% for the ATS49tg/E1KO7-5 *MboII*Δ *APRT* allele and 55% vs. 53% for the ATS49-Rv/E1KO-Rv *EcoRV*Δ *APRT* allele), but there were significant differences when comparing within wild-type and *ERCC1* deficient cell lines (35% for ATS49tg vs. 55% for ATS49-Rv, $p = 0.04$ and 28% for E1KO7-5 vs. 53% for E1KO-Rv, $p = 0.04$). The increased number of vector correction recombinants obtained from targeting the *EcoRV*Δ *APRT* allele relative to the *MboII*Δ *APRT* allele in wild-type cells can be explained mechanistically by considering that any conversion tracts extending beyond ~450-bp in the 3' direction from the repaired gap will copy the *MboII* deletion in the *MboII*Δ target allele, transferring it to the repaired vector, thus rendering it inactive and thereby eliminating it from selection. On the other hand, gap repair by recombination using the *EcoRV*Δ *APRT* target allele would require conversion tracts greater than ~1200-bp in order to result in inactivation of the gap repaired vector recombinant by conversion from the gene target. Thus, an inactivating mutation in the chromosomal target sequence that is close to the gap (e.g., the *MboII*Δ mutation) will have a higher probability of being transferred to the vector (i.e., requires a shorter conversion tract) than a mutation farther from the gap (e.g., the *EcoRV*Δ mutation), and will result in fewer successful vector correction events, as reflected in our data. In the case of targeting vector correction recombinants generated by pAG6ins0.9, as we reported previously [5], only wild-type cells were capable of generating this recombinant class as a significant proportion of the total number of *APRT* recombinants. This result confirms the requirement for ERCC1-XPF in efficient gap repair of vectors having ends blocked with lengthy nonhomologies. The few targeting vector recombinants obtained from E1KO-Rv cells (7/139) are therefore likely due to the increased distance of the *EcoRV*Δ *APRT* inactivating mutation from the DSB in the plasmid, resulting in a higher probability of the successful generation of selectable *APRT*⁺ recombinants.

In addition to describing the effects of changing the distance of the inactivating mutation in the target alleles relative to the DSB in the targeting vector and *ERCC1* deficiency on gene targeting, another goal of our study was to address the mechanism of generating targeted insertions, and aberrant recombinants in *ERCC1* deficient cells. Changing the target *APRT*

allele from *MboII*Δ to *EcoRV*Δ not only allowed us to assess the effects of *ERCC1* deficiency on the requirements for effective target gene conversion, but also permitted examination of the effects of reversing “targeting polarity,” as the *EcoRV*Δ mutation is upstream from the DSB in the vector, instead of downstream as for the *MboII*Δ allele. As illustrated in Figures 2 and 3, there are three possible targeted insertion structures (i.e., +/+, +/-, and -/+). In our previous experiments using ATS49tg and E1KO7-5 cell lines with the exon 5 *MboII*Δ mutation [5], all the targeted insertions recovered were either +/+ or +/- recombinants, as shown in Table 2. The +/- structure is consistent with integration of the targeting vector by two-ended strand invasion and appropriate resolution of Holliday junctions to result in crossover-type recombination products. In yeast [37] and mammalian cells [38] targeted gene replacement proceeds primarily by two-ended invasion, therefore this mechanism is likely responsible for generating targeted insertion recombinants in our system. The +/+ targeted insertions could arise either from conversion of the target by a dimerized targeting vector, or by integration (crossover) and co-conversion of the mutation; however, these mechanisms can not be rigorously distinguished by our analyses. Results in Table 2 show that *ERCC1* deficiency resulted in statistically significant decreases in the proportions of +/+ recombinants generated in both pairs of cell lines targeted with pAG6ins0.9 (i.e., 9% in ATS49tg vs. 2% in E1KO7-5, $p = 0.05$, and 18% in ATS49-Rv vs. 3% in E1KO-Rv, $p = 0.001$) but not when targeted with pAG6. Since conversion by a dimerized targeting vector would not be predicted to be dependent on the removal of blocked ends or *ERCC1*-XPF function, this result suggests that the +/+ subclass arise mainly through integration (resulting in a crossover-type recombinant) accompanied by co-conversion. Thus the reduction in +/+ recombinants in *ERCC1* deficient cells targeted with pAG6ins0.9 likely indicates defective resolution of intermediates leading to accompanying co-conversions in the targeted integration pathway. A similar recombination defective phenotype was observed in intrachromosomal mitotic recombination in *ERCC1* deficient cells by Sargent et al. [7]. In this case, “pop-out” crossover-type recombinants fell into two categories: simple crossovers and crossovers accompanied by conversion of the 5' adjacent *GPT* cassette. In wild-type cells, these recombinants were recovered in a 3:1 ratio, respectively, while in *ERCC1* deficient cells, the ratio was altered to 23:1. Results from studies of Al-Minawi et al. [33] and Talbert et al. [34] using DSB-induced intrachromosomal recombination reporters also support a function for *ERCC1*-XPF in resolution of recombination intermediates into conversion as well as crossover products. The results reported here therefore provide additional evidence that *ERCC1*-XPF is required to generate some types of gene conversion recombinants.

A striking result of reversing targeting polarity is evident by inspecting the distribution of subclasses of targeted insertion recombinants (see Fig. 2) in the different pairs of cell lines, as shown in Table 2. If the mechanism for generating targeted insertions proceeds through integration and crossover-type resolution of intermediates after two-ended strand invasion, as discussed above, then a logical prediction is that reversing the “targeting polarity” by altering the position of the target gene mutation relative to the DSB in the targeting vector will reverse the order of wild-type and mutant alleles in the resulting tandem *APRT* heteroallelic duplication. This is exactly what is observed in both wild-type cell lines; no +/- targeted insertion recombinants were recovered from targeting the exon 2 *EcoRV*Δ *APRT* allele and no -/+ insertion recombinants were recovered from targeting the exon 5 *MboII*Δ allele. In our previous study [5], the *ERCC1* deficient cell line E1KO7-5 was severely defective in generating targeted insertions from pAG6ins0.9, and this defective recombination phenotype is also reflected in the results of targeting E1KO-Rv with this vector (Table 2), confirming the requirement for *ERCC1*-XPF in generating crossover-type recombinants if the targeting vector has lengthy end-blocking nonhomologies.

Moreover, the aberrant recombinant class was recovered only from *ERCC1* deficient cells targeted with pAG6ins0.9. The structures of these recombinants resembled targeted insertions, except that variable-sized deletions extended downstream from the active *APRT* copy (see bottom structures in Fig. 2). We previously speculated [5] that these aberrant recombinants represented “aborted” vector integration events in which arrested strand-invasion recombination intermediates with unremoved nonhomologous ssDNA tails were encountered by replication forks, or otherwise experienced DSBs, and were then resolved through NHEJ pathways resulting in deletions. In the present study, we exploited the reversed targeting polarity afforded by the *EcoRV*Δ *APRT* allele in E1KO-Rv cells to test whether the mechanism for generating these aberrant recombinants was consistent with this model. As shown in Figure 4, the structures of the aberrant recombinants recovered from experiments in which E1KO-Rv was targeted with pAG6ins0.9 clearly reflect polarity in the “sidedness” of the deletions, consistent with our expected results based on the distribution of +/– and –/+ targeted insertion recombinants in the wild-type cell lines, as discussed above and illustrated in Figure 2. Targeting the *MboII*Δ *APRT* mutation (downstream from the DSB) resulted in recovery of only +/– targeted insertion recombinants and +/Δ aberrant recombinants (Fig. 2 bottom left) [5]. Reversing the targeting polarity by targeting the *EcoRV*Δ *APRT* mutation (upstream from the DSB,) resulted in the generation of only –/+ targeted insertion recombinants and aberrant Δ/+ recombinants with deletions in the upstream *APRT* allele (Fig. 2 bottom right; and see Fig. 4). This congruency supports a model in which aberrant recombinants recovered in the absence of ERCC1-XPF are generated through the same mechanism as targeted insertions, and likely arise from arrested strand-invasion recombination intermediates containing unremoved nonhomologous single-stranded tails. A possible model for the generation of this class of recombinants is illustrated in Figure 5.

Results of Table 2 also show that a significantly larger proportion of aberrant recombinants was generated from targeting the *EcoRV*Δ *APRT* allele with pAG6ins0.9 compared to the *MboII*Δ allele (48% vs. 16%, $p = 0.001$). Since there was a significant decrease depending on the target allele for target gene conversions (but not for targeting vector corrections) in wild-type cells targeted with this vector (46% for ATS49tg vs. 26% for ATS49-Rv, $p = 0.003$), it is likely that the requirement for a longer conversion tract from the vector DSB to convert the *EcoRV*Δ *APRT* allele, resulting in less productive target gene conversion, provided a relative excess of recombination intermediates for the strand-invasion pathway leading to plasmid integration and resulting in crossover-type targeted insertions. Therefore, if aberrant recombinants were generated through this pathway, the target allele-dependent increase we observed for aberrant recombinants in E1KO-Rv cells, dependent on loss of ERCC1-XPF functionality, as well as the distribution of +/– and –/+ targeted insertion subclasses, would be expected.

Studies of the role of ERCC1-XPF in homologous recombination have provided evidence that this structure-specific endonuclease performs important functions in resolving both single-strand annealing and strand-invasion recombination intermediates [5–7,33–35]. Results of the present study confirm and reinforce the requirement for ERCC1-XPF for removal of lengthy stretches of nonhomologous DNA sequence blocking the extension of 3'-OH homologous ends in gene targeting recombination intermediates. In addition, our results confirm the conclusions of other, recent studies [33,34] showing that ERCC1-XPF has important effects on gene conversion as well as crossover-type recombination, and is required for efficient co-conversion during targeted integration. Results of this study considered together with those of our previous study [5] also provide evidence supporting two-ended invasion and joint resolution to result in targeting vector integration as the primary mechanism for generating crossover-type recombinants in mammalian cell gene targeting. Furthermore, the generation of aberrant recombinants with the structures shown in

Figure 4, and in our previous study [5], strongly suggest that these recombinants result directly from misprocessing recombination intermediates through the targeted integration pathway in the absence of ERCC1-XPF.

Defective homologous recombination is a hallmark of many types of cancer, as observed in a number of genetic syndromes [39], and determining the mechanism(s) by which loss of ERCC1-XPF gives rise to anomalous recombination products increases our basic understanding of genomic instability and its relationship to cancer susceptibility. The results of this study provide further insights into the mechanisms of ERCC1-XPF in recombination, and directly confirm our previous speculation regarding the formation of aberrant recombinant structures with deletions, indicating that although alternate pathways exist for processing intermediates in cases where ERCC1-XPF is absent, these alternate pathways are error-prone and reflect increased genomic instability which increase the risk of deleterious mutations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Christine Brown for artwork and Becky Brooks for assistance in manuscript preparation. This work was supported by P01 award CA097175 from the National Cancer Institute and NIEHS Center of Excellence award ES007784. J.J.R. was supported by training grant T32 CA009480.

References

1. Aboussekhra A, Biggerstaff M, Shivji MK, Vilpo JA, Moncollin V, Podust VN, Protic M, Hubscher U, Egly JM, Wood RD. Mammalian DNA nucleotide excision repair reconstituted with purified protein components. *Cell*. 1995; 80:859–868. [PubMed: 7697716]
2. Costa RM, Chigancas V, Galhardo Rda S, Carvalho H, Menck CF. The eukaryotic nucleotide excision repair pathway. *Biochimie*. 2003; 85:1083–1099. [PubMed: 14726015]
3. Mu D, Hsu DS, Sancar A. Reaction mechanism of human DNA repair excision nuclease. *J Biol Chem*. 1996; 271:8285–8294. [PubMed: 8626523]
4. Wood RD. Nucleotide excision repair in mammalian cells. *J Biol Chem*. 1997; 272:23465–23468. [PubMed: 9295277]
5. Adair GM, Rolig RL, Moore-Faver D, Zabelshansky M, Wilson JH, Nairn RS. Role of *ERCC1* in removal of long non-homologous tails during targeted homologous recombination. *EMBO J*. 2000; 19:5552–5561. [PubMed: 11032822]
6. Sargent RG, Meservy JL, Perkins BD, Kilburn AE, Intody Z, Adair GM, Nairn RS, Wilson JH. Role of the nucleotide excision repair gene *ERCC1* in formation of recombination-dependent rearrangements in mammalian cells. *Nucleic Acids Res*. 2000; 28:3771–3778. [PubMed: 11000269]
7. Sargent RG, Rolig RL, Kilburn AE, Adair GM, Wilson JH, Nairn RS. Recombination-dependent deletion formation in mammalian cells deficient in the nucleotide excision repair gene *ERCC1*. *Proc Natl Acad Sci U S A*. 1997; 94:13122–13127. [PubMed: 9371810]
8. Wu Y, Mitchell TR, Zhu XD. Human XPF controls TRF2 and telomere length maintenance through distinctive mechanisms. *Mech Ageing Dev*. 2008; 129:602–610. [PubMed: 18812185]
9. Zhu XD, Niedernhofer L, Kuster B, Mann M, Hoeijmakers JH, de Lange T. ERCC1/XPF removes the 3' overhang from uncapped telomeres and represses formation of telomeric DNA-containing double minute chromosomes. *Mol Cell*. 2003; 12:1489–1498. [PubMed: 14690602]
10. Davies AA, Friedberg EC, Tomkinson AE, Wood RD, West SC. Role of the Rad1 and Rad10 proteins in nucleotide excision repair and recombination. *J Biol Chem*. 1995; 270:24638–24641. [PubMed: 7559571]

11. de Laat WL, Jaspers NG, Hoeijmakers JH. Molecular mechanism of nucleotide excision repair. *Genes Dev.* 1999; 13:768–785. [PubMed: 10197977]
12. Tomkinson AE, Bardwell AJ, Bardwell L, Tappe NJ, Friedberg EC. Yeast DNA repair and recombination proteins Rad1 and Rad10 constitute a single-stranded-DNA endonuclease. *Nature.* 1993; 362:860–862. [PubMed: 8479526]
13. Bessho T, Sancar A, Thompson LH, Thelen MP. Reconstitution of human excision nuclease with recombinant XPF-ERCC1 complex. *J Biol Chem.* 1997; 272:3833–3837. [PubMed: 9013642]
14. de Laat WL, Appeldoorn E, Jaspers NG, Hoeijmakers JH. DNA structural elements required for ERCC1-XPF endonuclease activity. *J Biol Chem.* 1998; 273:7835–7842. [PubMed: 9525876]
15. Evans E, Moggs JG, Hwang JR, Egly JM, Wood RD. Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *EMBO J.* 1997; 16:6559–6573. [PubMed: 9351836]
16. Matsunaga T, Park CH, Bessho T, Mu D, Sancar A. Replication protein A confers structure-specific endonuclease activities to the XPF-ERCC1 and XPG subunits of human DNA repair excision nuclease. *J Biol Chem.* 1996; 271:11047–11050. [PubMed: 8626644]
17. Colaiacovo MP, Paques F, Haber JE. Removal of one nonhomologous DNA end during gene conversion by a *RAD1*- and *MSH2*-independent pathway. *Genetics.* 1999; 151:1409–1423. [PubMed: 10101166]
18. Ivanov EL, Haber JE. *RAD1* and *RAD10*, but not other excision repair genes, are required for double-strand break-induced recombination in *Saccharomyces cerevisiae*. *Mol Cell Biol.* 1995; 15:2245–2251. [PubMed: 7891718]
19. Prado F, Aguilera A. Role of reciprocal exchange, one-ended invasion crossover and single-strand annealing on inverted and direct repeat recombination in yeast: different requirements for the *RAD1*, *RAD10*, and *RAD52* genes. *Genetics.* 1995; 139:109–123. [PubMed: 7705617]
20. Bardwell AJ, Bardwell L, Tomkinson AE, Friedberg EC. Specific cleavage of model recombination and repair intermediates by the yeast Rad1-Rad10 DNA endonuclease. *Science.* 1994; 265:2082–2085. [PubMed: 8091230]
21. Aguilera A, Klein HL. Yeast intrachromosomal recombination: long gene conversion tracts are preferentially associated with reciprocal exchange and require the *RAD1* and *RAD3* gene products. *Genetics.* 1989; 123:683–694. [PubMed: 2558957]
22. Schiestl RH, Prakash S. *RAD1*, an excision repair gene of *Saccharomyces cerevisiae*, is also involved in recombination. *Mol Cell Biol.* 1988; 8:3619–3626. [PubMed: 3065620]
23. Pennington SL, Wilson JH. Gene targeting in Chinese hamster ovary cells is conservative. *Proc Natl Acad Sci U S A.* 1991; 88:9498–9502. [PubMed: 1946364]
24. Tindall KR, Stankowski LF Jr, Machanoff R, Hsie AW. Detection of deletion mutations in pSV2gpt-transformed cells. *Mol Cell Biol.* 1984; 4:1411–1415. [PubMed: 6095070]
25. Adair GM, Stallings RL, Nairn RS, Siciliano MJ. 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 U S A.* 1983; 80:5961–5964. [PubMed: 6310607]
26. Smith DG, Adair GM. Characterization of an apparent hotspot for spontaneous mutation in exon 5 of the Chinese hamster *APRT* gene. *Mutat Res.* 1996; 352:87–96. [PubMed: 8676921]
27. Clarkson JM, Mitchell DL, Adair GM. The use of an immunological probe to measure the kinetics of DNA repair in normal and UV-sensitive mammalian cell lines. *Mutat Res.* 1983; 112:287–299. [PubMed: 6633556]
28. Rolig RL, Lowery MP, Adair GM, Nairn RS. Characterization and analysis of Chinese hamster ovary cell *ERCC1* mutant alleles. *Mutagenesis.* 1998; 13:357–365. [PubMed: 9717172]
29. Rolig RL, Layher SK, Santi B, Adair GM, Gu F, Rainbow AJ, Nairn RS. Survival, mutagenesis, and host cell reactivation in a Chinese hamster ovary cell *ERCC1* knock-out mutant. *Mutagenesis.* 1997; 12:277–283. [PubMed: 9237774]
30. Limoli CL, Kaplan MI, Phillips JW, Adair GM, Morgan WF. Differential induction of chromosomal instability by DNA strand-breaking agents. *Cancer Res.* 1997; 57:4048–4056. [PubMed: 9307292]
31. Nairn RS, Adair GM. Use of gene targeting to study recombination in mammalian cell DNA repair mutants. *Methods Mol Biol.* 2006; 314:133–154. [PubMed: 16673880]

32. Adair GM, Nairn RS, Wilson JH, Seidman MM, Brothman KA, MacKinnon C, Scheerer JB. Targeted homologous recombination at the endogenous adenine phosphoribosyltransferase locus in Chinese hamster cells. *Proc Natl Acad Sci U S A*. 1989; 86:4574–4578. [PubMed: 2734308]
33. Al-Minawi AZ, Saleh-Gohari N, Helleday T. The ERCC1/XPF endonuclease is required for efficient single-strand annealing and gene conversion in mammalian cells. *Nucleic Acids Res*. 2008; 36:1–9. [PubMed: 17962301]
34. Talbert LL, Coletta LD, Lowery MG, Bolt A, Trono D, Adair GM, Nairn RS. Characterization of CHO *XPF* mutant UV41: influence of *XPF* heterozygosity on double-strand break-induced intrachromosomal recombination. *DNA Repair (Amst)*. 2008; 7:1319–1329. [PubMed: 18547876]
35. Niedernhofer LJ, Essers J, Weeda G, Beverloo B, de Wit J, Muijtjens M, Odijk H, Hoeijmakers JH, Kanaar R. The structure-specific endonuclease Ercc1-Xpf is required for targeted gene replacement in embryonic stem cells. *EMBO J*. 2001; 20:6540–6549. [PubMed: 11707424]
36. Symington LS, Kang LE, Moreau S. Alteration of gene conversion tract length and associated crossing over during plasmid gap repair in nuclease-deficient strains of *Saccharomyces cerevisiae*. *Nucleic Acids Res*. 2000; 28:4649–4656. [PubMed: 11095674]
37. Langston LD, Symington LS. Gene targeting in yeast is initiated by two independent strand invasions. *Proc Natl Acad Sci U S A*. 2004; 101:15392–15397. [PubMed: 15489271]
38. Li J, Read LR, Baker MD. The mechanism of mammalian gene replacement is consistent with the formation of long regions of heteroduplex DNA associated with two crossing-over events. *Mol Cell Biol*. 2001; 21:501–510. [PubMed: 11134338]
39. Hoeijmakers JH. Genome maintenance mechanisms for preventing cancer. *Nature*. 2001; 411:366–374. [PubMed: 11357144]

Figure 1. Targeted recombination at the CHO *APRT* locus

A. Each targeting vector carries a full-length copy of the Chinese hamster *APRT* gene disrupted by insertions of 33-bp (pAG6) or 945-bp (pAG6ins0.9) of nonhomologous sequence. Digestion of targeting vector pAG6 with *Hind*III yields a linear ends-in insertion vector configuration in which 18- and 11-nt terminal nonhomologies flank each side of the double-strand break after resection. Digestion of targeting vector pAG6ins0.9 with *Sal*I yields a similar linearized vector with 271- and 670-nt terminal nonhomologies flanking the double-strand break after resection. Both targeting vectors contain a *GPT* cassette in the plasmid backbone, which is used to measure non-targeted (i.e., random) integration events. The cell lines used in this study (ATS49-Rv and E1KO-Rv) have the same 2-bp deletion in exon 2 of the *APRT* target gene locus resulting in loss of an *EcoRV* restriction site (*EcoRV*Δ). **B.** Electroporation of these cell lines with either targeting vector, followed by selection for ALASA-resistant clones, results in recovery of *APRT*⁺ recombinants with the recombinant class structures shown: GC, target gene conversion arising by unidirectional transfer of the wild-type *APRT* sequence to the chromosomal locus restoring functional *APRT*; TI, targeted insertion arising by single-crossover reciprocal exchange whereby the targeting vector is integrated into the target gene locus; VC, vector correction arising by gap repair of the targeting vector and subsequent integration of the vector elsewhere in the cellular genome. B, *Bam*HI; Bg, *Bgl*II; H, *Hind*III; RV, *EcoRV*; S, *Sal*I; X, *Xho*I.

Figure 2. Design of gene targeting experiments using different gene target mutant alleles to investigate the mechanism of generating targeted insertions and aberrant recombinants

This Figure illustrates the experimental strategy used to address the mechanism of formation of targeted insertions and aberrant recombinants. The targeting vector shown is pAG6ins0.9, which unlike pAG6, produces aberrant recombinants in *ERCC1* deficient cells, and also results in *ERCC1* dependent ablation of targeted insertion recombinants [5]. In both wild-type and *ERCC1* deficient cells, pAG6ins0.9 targeting of either *APRT* allele (i.e., *Mboll*Δ, left side of diagram, used in our previous study [5], or *EcoRV*Δ, right side, used in this study) generates target gene conversion recombinants, although in different proportions depending on the target allele (see text). However, targeted insertion recombinants are generated by pAG6ins0.9 only in wild-type cells, and aberrant recombinants are only generated in *ERCC1* deficient cells. The structures of targeted insertion recombinants with characteristics of arising from a single crossover (+/-) from our previous study (left side; see [5]) invariably retained the original target *APRT* allele downstream in the resulting heteroallelic duplication, and aberrant recombinants from this study had deletions of variable size in the downstream *APRT* copy, but never in the upstream copy (+/Δ). By targeting the *EcoRV*Δ *APRT* allele (right side of diagram) in cell lines ATS-Rv and E1KO-Rv in the present study, we tested whether reversing the “targeting polarity” by altering the position of the target gene mutation from downstream to upstream from the DSB in pAG6ins0.9 would confirm the apparent mechanism of generating targeted insertions by single-reciprocal exchange in wild-type cells and additionally test the hypothesis that aberrant recombinants in *ERCC1* deficient cells were produced through the same pathway. The predictions and results of this experimental design are discussed in the text.

Figure 3. Determination of *APRT* recombinant structures

DNA samples from recombinant clones were co-digested with pairs of restriction enzymes as shown and analyzed by Southern blotting to determine the recombinant class for each. The *APRT* probe is shown as a gray bar in **A**. **A. – E.** Diagrams of recombinant class structures with sizes of the restriction fragments generated for each shown beneath each structure. The class type of each recombinant structure is indicated, i.e. “target gene conversion”, etc. The structures of different targeted insertion subclasses (**B. – D.**) are shown together with their diagnostic restriction enzyme digestion fragments. As indicated at the right margin, *BglII/XhoI* digests were used to distinguish target gene conversions (GC) from targeted insertions (TI) and target vector corrections (VC). *EcoRV/HindIII* digests were informative in discriminating between GC and TI or VC events and were also used to distinguish the three different subclasses of TI events (TI +/-, TI +/+, TI -/+). VC events are unique in that they have the bands corresponding to the unchanged target locus in addition to extra bands that vary in size and intensity from recombinant to recombinant. Aberrant recombinants (not shown) had restriction patterns where certain fragments were smaller in size or missing altogether (also see Figure 4 and Supplemental Figure 3). Supplemental Figures 1 and 2 show representative Southern blots for the recombinant structures illustrated in this Figure. Asterisks (*) indicate fragments that are light in intensity on Southern blots due to less homology between the probe and the genomic fragment.

Figure 4. Structures of representative aberrant recombinants recovered from E1KO-Rv cells (*EcoRV*Δ target allele) targeted with pAG6ins0.9

The structures of these seven aberrant events were deduced from the Southern blots shown in Supplemental Figure 3. The clone names are derived from the cell line designation and experiment. See Figure 3 and Supplemental Figure 1 and their legends, and Supplemental Figure 3, for an explanation of how these structures were deduced.

Figure 5. Model for generating aberrant recombinants in *ERCC1* deficient cells

a. Targeting vector pAG6ins0.9 is linearized and transfected into cells, and after the resection of the DNA ends, **b.**, long ssDNA "tails" are exposed in which homologous *APRT* DNA sequence (dark grey) is blocked by heterologous DNA sequence (black). Strand invasion by the resected ends into the genomic *APRT* target locus (light grey), **c.**, generates recombination intermediates which are further processed in the presence (**d. – f.**) or absence (**g. – j.**) of ERCC1-XPF in wild-type and *ERCC1* deficient CHO cell lines, respectively. As shown in the diagram, the vertical line in the genomic *APRT* sequence represents the *EcoRV* deletion mutation. In wild-type cells, the end-blocking heterologous DNA sequence (black) is incised by ERCC1-XPF, (**d.**), exposing a free 3'-OH homologous end which can be extended using genomic DNA sequence as template (**e.**). This results in copying of the exon 2 *EcoRV* deletion mutation, and the retention of this mutation in the upstream copy of the recombination product (**f.**). In *ERCC1* deficient cells, the end-blocking sequence can not be removed by ERCC1-XPF to allow extension by DNA synthesis (**g.**). If there is break formation (**h.**), then a non-homology dependent process may result in joining of the free

ends with accompanying formation of a deletion of variable size (**i.**), followed by extension and resolution into a recombination product with the upstream copy showing evidence of this deletogenic process (**j.**).

Table 1

Frequencies of homologous recombination in *ERCC1*⁺ and *ERCC1*[−] cell lines^a after electroporation with linearized pAG6 or pAG6ins0.9 targeting vectors

Targeting vector/cell line	Number of experiments	Frequency of APRT ⁺ recombinants	Frequency of GPT ⁺ recombinants	Ratio APRT ⁺ /GPT ⁺
<i>HindIII</i> -cut pAG6				
ATS49tg (WT)	5	$5.5 \pm 0.7 \times 10^{-6}$	$13.7 \pm 2.9 \times 10^{-4}$	1/249
E1KO7-5 (<i>ERCC1</i> ^{ko})	2	$4.3 \pm 1.4 \times 10^{-6}$	$12.6 \pm 4.2 \times 10^{-4}$	1/293
ATS49-Rv (WT)	4	$11.8 \pm 2.2 \times 10^{-6}$	$53.2 \pm 7.2 \times 10^{-4}$	1/451
E1KO-Rv (<i>ERCC1</i> ^{ko})	4	$6.0 \pm 1.2 \times 10^{-6}$	$28.7 \pm 6.2 \times 10^{-4}$	1/478
<i>Sall</i> -cut pAG6ins0.9				
ATS49tg (WT)	14	$5.1 \pm 0.9 \times 10^{-6}$	$10.9 \pm 1.0 \times 10^{-4}$	1/214
E1KO7-5 (<i>ERCC1</i> ^{ko})	10	$3.0 \pm 0.7 \times 10^{-6}$	$8.8 \pm 1.7 \times 10^{-4}$	1/293
ATS49-Rv (WT)	6	$4.6 \pm 0.6 \times 10^{-6}$	$50.3 \pm 5.4 \times 10^{-4}$	1/1094
E1KO-Rv (<i>ERCC1</i> ^{ko})	6	$3.4 \pm 0.5 \times 10^{-6}$	$35.4 \pm 3.4 \times 10^{-4}$	1/1041

^aData from cell lines ATS49tg and E1KO7-5 are from a previously published study [5] and are included for comparison; see text.

Table 2

Class distributions of targeted recombination events in *ERCC1*⁺ and *ERCC1*⁻ cell lines^a after electroporation with linearized pAG6 or pAG6ins0.9 targeting vectors

Targeting vector/cell line	APRT target allele	APRT ⁺ recombinants				
		Target gene conversions	Targeting vector corrections	Targeted insertions		
				+/-	+/+	-/+
<i>HindIII</i> -cut pAG6						
ATS49tg (WT)	<i>MboII</i> Δ	17/49 (35%)	17/49 (35%)	10/49 (20%)	5/49 (10%)	0/49 (0%)
E1KO7-5 (<i>ERCC1</i> ^{ko})	<i>MboII</i> Δ	12/29 (41%)	8/29 (28%)	3/29 (10%)	6/29 (21%)	0/29 (0%)
ATS49-Rv (WT)	<i>EcoRV</i> Δ	6/38 (16%)	21/38 (55%)	0/38 (0%)	1/38 (3%)	10/38 (26%)
E1KO-Rv (<i>ERCC1</i> ^{ko})	<i>EcoRV</i> Δ	6/36 (17%)	19/36 (53%)	0/36 (0%)	2/36 (6%)	9/36 (25%)
<i>SacI</i> -cut pAG6ins0.9						
ATS49tg (WT)	<i>MboII</i> Δ	41/90 (46%)	14/90 (16%)	27/90 (30%)	8/90 (9%)	0/90 (0%)
E1KO7-5 (<i>ERCC1</i> ^{ko})	<i>MboII</i> Δ	55/67 (82%)	0/67 (0%)	0/67 (0%)	1/67 (2%)	0/67 (0%)
ATS49-Rv (WT)	<i>EcoRV</i> Δ	26/101 (26%)	23/101 (23%)	0/101 (0%)	18/101 (18%)	42/101 (42%)
E1KO-Rv (<i>ERCC1</i> ^{ko})	<i>EcoRV</i> Δ	53/139 (38%)	7/139 (5%)	0/139 (0%)	4/139 (3%)	9/139 (7%)

^aData from cell lines ATS49tg and E1KO7-5 are from a previously published study [5] and are included for comparison; see text.