Human Gene Targeting Favors Insertions Over Deletions

David W. Russell^{1,2} and Roli K. Hirata¹

Abstract

Gene targeting is a powerful technique for manipulating the human genome, but few studies have directly compared the targeting frequencies of various types of vector constructs. Here we show that similar targeting constructs are able to insert nucleotides at the homologous chromosomal target locus more efficiently than they can delete nucleotides, and combination insertion/deletion vectors appear to target at intermediate frequencies. This holds true for deletions ranging from 1 to 334 bp and insertions ranging from 1 to 1332 bp. In addition, vectors designed to inactivate the human hypoxanthine phosphoribosyltransferase gene (*HPRT*) by deleting nucleotides often produced rearrangements at the target locus that in many cases were due to insertions of multimerized vector constructs, effectively converting a deletion vector into an insertion vector. These findings were obtained when adeno-associated virus vectors were used to efficiently deliver single-stranded DNA targeting constructs, but the same phenomenon was also observed when transfecting linearized double-stranded plasmids. Thus human cells distinguish between deletion and insertion vectors and process their recombination intermediates differently, presumably at the heteroduplex stage, with implications for the design of gene-targeting vectors and the evolution of human genomes.

Introduction

GENE TARGETING can be used to create mutations at specific chromosomal sequences by homologous recombination. In mammalian cells, several types of mutations have been introduced by transfecting or electroporating targeting constructs, including deletions, insertions, and point mutations. Typically, targeting constructs are plasmids linearized within the homology arms that introduce a selectable marker and other genetic elements at the chromosomal target locus. If necessary, subsequent recombination steps can then convert these insertions into more subtle mutations or deletions. Because the initial gene-targeting reaction creates an insertion, and there are powerful selection strategies to improve the recovery of recombinants, most studies do not address the relative targeting frequencies of different types of constructs. However, there are still cases in which improved targeting frequencies can be important, such as the genetic manipulation of recalcitrant cell types, manipulations at unexpressed loci, or therapeutic gene targeting, and thus variations in vector design could significantly impact experimental outcomes.

In the case of the *hypoxanthine phosphoribosyltransferase* (*HPRT*) locus, the targeting frequencies of various constructs can be easily compared, because cells with mutations in this single-copy, X-linked gene are resistant to 6-thioguanine (6TG), and *HPRT*⁺ cells can be grown in HAT medium (con-

tains hypoxanthine, aminopterin, and thymidine), allowing one to directly select for a variety of different genetic modifications. This system has been widely used to measure the targeting frequencies of various plasmid constructs (Thomas and Capecchi, 1986; Hasty *et al.*, 1991; Valancius and Smithies, 1991; Zhang *et al.*, 1994). These types of studies have not reported dramatic differences in the targeting rates of vectors designed to create insertions versus those designed to create deletions.

Vectors based on adeno-associated virus (AAV), which deliver single-stranded, linear DNA genomes, are able to efficiently introduce many types of mutations into homologous target loci in mammalian cells (Hendrie and Russell, 2005; Vasileva and Jessberger, 2005). Gene targeting can occur in as much as 1% of the entire cell population exposed to AAV vectors (Russell and Hirata, 1998; Hirata *et al.*, 2002), making this approach particularly useful for manipulating primary human cells with therapeutic potential (Chamberlain *et al.*, 2004, 2008). We noted that the targeting frequencies of different AAV vectors can vary significantly depending on the type of mutation being introduced (Inoue *et al.*, 1999, 2001; Hirata and Russell, 2000; Hirata *et al.*, 2002), prompting us to further evaluate this phenomenon. Here we have compared the targeting frequencies of AAV vectors designed to introduce insertion and deletion mutations, and found that insertions are consistently introduced at higher frequencies. Analogous experiments with transfected plasmid

¹Department of Medicine and ²Department of Biochemistry, University of Washington, Seattle, WA 98195.

constructs resulted in a similar preference for insertions, suggesting that it is a general feature of human gene targeting and recombination.

Materials and Methods

Cell culture

HT-1080 male human sarcoma cells (Rasheed *et al.*, 1974), MHF2 normal male human fibroblasts (GM05387; Coriell Institute for Medical Research, Camden, NJ), Tc7 human fibroblasts immortalized by the human telomerase reverse transcriptase gene (*hTERT*) (Hirata *et al.*, 2002), and PG13 retroviral vector packaging cells (Miller *et al.*, 1991) were cultured in Dulbecco's modified Eagle's medium (DMEM) with 10% heat-inactivated (56°C for 30 min) fetal bovine serum (HyClone, Logan, UT), amphotericin (1.25 µg/ml), penicillin (100 U/ml), and streptomycin (100 µg/ml) at 37°C in a 5% CO₂ atmosphere. HT-1080 and MHF2 cells were maintained in HAT medium (DMEM containing hypoxanthine [13.6 µg/ml], aminopterin [0.18 µg/ml], and thymidine [3.88 µg/ml]) before their use in *HPRT* gene-targeting experiments that inactivated *HPRT*.

Plasmids, vectors, and DNA analysis

Plasmids pA2HPe3X (Hirata *et al.*, 2002), pA2HPe3PN (Hirata *et al.*, 2002), pA2-5' AP-Bss (Hirata and Russell, 2000), and pLAPSN (Miller *et al.*, 1994) have been described. Plasmids pA2HPe3(Δ4) and pA2HPe3 are identical to pA2HPe3X except that the 4-bp *HPRT* exon 3 insertion in pA2HPe3X was replaced with a 4-bp deletion or wild-type (WT) sequence, respectively. Plasmids pA2HPe3X and pA2HPe3(Δ4) both destroy a *Bso*BI site present in the wild-type exon 3. Plasmids pA2HPΔe3 and pA2HPΔe3PN are identical to plasmids pA2HPe3 and pA2HPe3PN, respectively, but contain a 334-bp deletion from a *Bsa*AI site to the *Xho*I site in exon 3 of *HPRT*. Plasmid pLAPIH was made by replacing the SV40-*neo* cassette of pLAPSN with an internal ribosome entry site (IRES) and hygromycin resistance gene. Plasmids pLAP375(+1)IH and pLAP375(Δ1)IH are identical to pLAPIH, but contain a 1-bp insertion or deletion, respectively, at bp 375 of the alkaline phosphatase (AP) reading frame. Plasmid pHPe2/3 contains an *Eco*RI fragment of λ phage Huλ3 (Patel *et al.*, 1986). All plasmid sequences are available on request.

AAV vector stocks were prepared by transient cotransfection of helper and vector plasmids and purification on density gradients, and then titered by Southern blots to determine the number of genome-containing particles as described (Hirata *et al.*, 2002). All stocks were serotype 2. Plasmids pA2HPe3, pA2HPe3X, pA2HPe3(Δ4), pA2HPΔe3, pA2HPΔe3PN, pA2HPe3PN, and pA2-5' AP-Bss were used to generate vectors AAV-HPe3, AAV-HPe3(+4), AAV-HPe3(Δ4), AAVHPe3(Δ334), AAV-HPe3PN(Δ334), AAV-HPe3PN, and AAV-5' APBss, respectively. Retroviral vector stocks pseudotyped with the gibbon ape leukemia virus envelope were prepared in Phoenix amphotropic packaging cells as described (http://www.stanford.edu/group/nolan/protocols/pro_helper_dep.html). Plasmids pLAP375(+1)IH and pLAP375(Δ1)IH were used to make vectors MLV-LAP375(+1)IH and MLV-LAP375(Δ1)IH, respectively.

Genomic DNA was isolated according to the Puregene DNA purification protocol (Gentra Systems/Qiagen, Min-

neapolis, MN), and analyzed by Southern blots according to standard protocols. The probe fragment shown in Figs. 1 and 2 (see below) spans nucleotides 133436892–133438049 of the X chromosome (March 2006 freeze of the human genome sequence).

AAV-mediated gene-targeting experiments

When knocking out *HPRT*, MHF2 cells maintained in HAT medium were plated at 5×10^4 cells per well in 24-well plates in DMEM on day 1, infected with AAV vectors on day 2 at a multiplicity of infection (MOI) of 20,000 genome-containing particles per cell, passaged with trypsin on day 3, plated in a 10-cm dish, and then expanded for 10 days in DMEM to eliminate existing *HPRT* protein. After this phenotypic expression period, 10^5 , 10^4 , and 10^2 cells from each culture were plated into new 10-cm dishes, and the following day 6TG (10 µg/ml) was added to the 10^5 and 10^4 cell dishes. The 10^2 cell dishes were used to calculate plating efficiencies. In some cases additional dishes were also plated at this time for G418 selection (active compound, 0.6 mg/ml). The cells were cultured for 10–13 additional days and the surviving colonies were counted. Experiments with HT-1080 cells were done in the same way, except that the initial plating was 2×10^4 cells per well in 48-well plates, and the AAV vector MOI was 50,000.

When correcting *HPRT*, HT-1080 subclones with *HPRT* mutations were cultured in DMEM, plated at 2×10^4 cells per well in 48-well plates on day 1, infected with AAV-HPe3 at an MOI of 50,000 on day 2, and passaged with trypsin on day 3 to plate dilutions of 89, 10, and 1% of cells in 10-cm dishes. On day 5, the 89 and 10% platings were placed in HAT medium and the 1% plating was cultured without selection. The surviving colonies were counted 7–10 days later.

To measure AP gene-targeting rates, we first generated polyclonal populations of Tc7 cells transduced with MLV-LAP375(+1)IH or MLV-LAP375(Δ1)IH essentially as described (Hirata and Russell, 2000), except that hygromycin selection was used (150 µg/ml). We used an MOI of <1 transducing unit/cell to minimize the number of cells containing two proviruses, and the hygromycin-resistant polyclonal populations were derived from $>10^4$ independent transduction events. For AP gene targeting, Tc7 cells containing MLV-LAP375(+1)IH or MLV-LAP375(Δ1)IH provirus were seeded on day 1 at 5×10^4 cells per well in 48-well plates, infected with AAV-5' APBss on day 2 at an MOI of 20,000, treated with trypsin, and plated in two 6-cm dishes on day 3 (0.25 or 99.75% of cells). The 0.25% plating was stained on day 4 and used to calculate the total number of viable cells per well. The 99.75% plating was cultured until day 9, and then stained histochemically for AP expression as described (Fields Berry *et al.*, 1992). Gene correction rates were calculated as the number of AP⁺ foci per 10^5 viable cells.

Transfection-targeting experiments

HT-1080 subclones with *HPRT* mutations were cultured in DMEM, plated at 5×10^5 cells per 10-cm dish (11 dishes for each subclone) on day 1, and transfected with pHPe2/3 (predigested with *Hind*III and *Eco*RI to produce the fragment shown in Fig. 4; see below) and pCMVβ (which expresses the *lacZ* gene from a cytomegalovirus [CMV] promoter; Clontech, Palo Alto, CA) as a transfection control on day 2

FIG. 1. Targeting *HPRT* with insertion and deletion vectors. (A) The structures of five AAV gene-targeting vectors and a section of the human chromosomal *HPRT* locus are shown with the viral inverted terminal repeats as hatched boxes, homology arms as gray boxes, and *HPRT* exons as black boxes; Δ , deletions. Locations of the PGK promoter; *neo* gene; *Hind*III (H), *Ahd*I (A), and *Bso*BI (B) restriction sites; and probe used in Southern blots are also shown. (B) The fraction of colonies resistant to 6TG (solid columns) or G418 (open columns) obtained after infection of normal human male fibroblasts (MHF2 cells) with the indicated AAV-HPe3-based vector is shown (mean $\pm$ standard deviation, n : 3). The hatched column represents the fraction of 6TG-resistant colonies obtained without infection (mean of six experiments in which 6TG-resistant colonies were obtained only once). (C) Southern blots of representative 6TG-resistant MHF2 clones transduced with AAV-HPe3PN, AAV-HPe3PN(Δ 334), or AAV-HPe3(Δ 334) as indicated. The left and right lanes represent parental, untransduced cells. Each lane contains genomic DNA digested with *Hind*III and *Ahd*I and probed as shown in (A). The positions of size standards are shown on the left. (D) As in (C) but done in HT-1080 cells.

(each dish received 10 μ g of pHPe2/3 and 0.1 μ g of pCMV β), using the Superfect reagent (Qiagen, Valencia, CA) and according to the manufacturer's recommended protocol. On day 3, the transfected cells were treated with trypsin and pooled, 5×10^4 cells were plated in a well of a 6-well plate for β -galactosidase staining, and the remaining cells were counted and plated in ten 15-cm dishes. On day 4, the 6-well plate was stained for β -galactosidase expression (Fields Berry *et al.*, 1992) and the number of positive foci was determined. On day 5, the remaining dishes were switched to HAT medium for selection, and the surviving colonies were counted 10–14 days later. Figure 4B (see below) shows the number of HAT-resistant colonies divided by the number of cells plated in HAT medium and normalized for differences in transfection efficiency based on staining for β -galactosidase. A negative control dish was treated in the same way for each subclone, but no plasmid was used in the transfection mixture, to determine the background reversion rate at *HPRT*.

Results

Deletion vectors target at lower frequencies

We designed several AAV targeting vectors to introduce inactivating mutations into exon 3 of the human *HPRT* gene (Fig. 1A). These vectors introduce either a relatively large insertion of 1.3 kb consisting of a neomycin phosphotransferase gene (*neo*) driven by a phosphoglycerate kinase (PGK) promoter (AAV-HPe3PN), a 334-bp deletion (AAV-HPe3(Δ 334)), a combination insertion/deletion (AAV-HPe3PN(Δ 334)), or a small, 4-bp insertion or deletion (AAV-HPe3(+4) or AAV-HPe3(Δ 4)). We infected normal, male human fibroblasts (MHF2 cells) with each of these vectors, plated for colony growth, and determined the fraction of colonies that became 6TG resistant through inactivation of the *HPRT* gene, which represented gene-targeting events. In the case of vectors AAV-HPe3PN and AAV-HPe3PN(Δ 334), we were also able to select for G418-resistant colonies that represented either targeted or random integration events.

TABLE 1. SUMMARY OF SOUTHERN BLOT RESULTS

AAV vector	Cell line	Number of 5TG-resistant colonies				
		Analyzed	Targeted ^a	Accurate ^b	Dimer ^c	Other bands ^d
HPe3(Δ 4)	MHF2	17	13	2	11	3
HPe3(Δ 4)	HT-1080	6	6	4	2	0
HPe3(+4)	MHF2	29	28	28	0	2
HPe3(+4)	HT-1080	12	12	12	0	1
HPe3(Δ 334)	MHF2	12	6	3	ND ^e	3
HPe3(Δ 334)	HT-1080	11	7	3	ND ^e	4
HPe3(Δ 334)PN	MHF2	17	17	16	0	1
HPe3(Δ 334)PN	HT-1080	6	6	6	0	0
HPe3PN	MHF2	6	6	6	0	0
HPe3PN	HT-1080	6	6	6	0	0

^aIndicates a change in the wild-type *HPRT* locus by Southern blots.

^bAll fragments seen on Southern blots consistent with accurate replacement of the wild-type locus with the mutation present in the targeting vector.

^cSouthern blots consistent with homologous recombination between a vector genome dimer and the wild-type locus, with duplication of the vector sequences at the locus.

^dThese bands could represent random integrants or other unexplained rearrangements at the locus.

^eSouthern blot pattern too complicated to characterize, but could represent recombination with concatemericized vector genomes.

Each vector produced 6TG-resistant colonies above the background *HPRT* mutation frequency of uninfected fibroblasts (Fig. 1B). The two deletion vectors AAV-HPe3(Δ 4) and AAVHPe3(Δ 334) had similar targeting frequencies, whereas the insertion and insertion/deletion vectors AAV-HPe3(+4), AAV-HPe3PN, and AAV-HPe3PN(Δ 334) all had higher targeting frequencies. In the case of the Δ 4 and +4 vectors there was a 5.7-fold difference, even though only 8 bases differed between the vector genomes. In the case of larger insertions and deletions, 6TG resistance was 14.9-fold greater for the PN vector as compared with the Δ 334 vector, and 2.4-fold greater than that of the combination insertion/deletion PN(Δ 334) vector. The PN and PN(Δ 334) vectors also carried a selectable *neo* gene, and the fraction of G418-resistant colonies was similar for both vectors (0.045 vs. 0.039), indicating that random integration was not affected by the vector structure, and serving as an internal control. One variable in these studies was the length of homology for each vector, because longer homology arms are known to increase targeting frequencies (Hirata and Russell, 2000). In the case of the two Δ 334 vectors, the 334-bp deletion removed 11% of the 2914-bp homology to the target locus, which may have influenced targeting. However, the Δ 4 vector had only 4 bp less homology than the +4 and PN vectors, yet targeting frequencies varied over 22-fold for these vectors. These findings all suggest that insertion vectors have higher targeting frequencies than deletion vectors, and that combined insertion/deletion vectors target at intermediate levels.

Deletion vectors target inaccurately

We expanded several 6TG-resistant colonies from the experiments described above, as well as additional colonies from HT-1080 human fibrosarcoma cells infected with the same AAV targeting vectors, and analyzed the *HPRT* locus by Southern blots. In cells infected with the +4, PN, or PN(Δ 334) vectors designed to create insertions or a combined insertion/deletion, almost every clone analyzed had under-

gone targeting at the *HPRT* locus (75 of 76 clones; Table 1), confirming that 6TG resistance was an accurate measure of targeting frequencies for these vectors. This was not the case for the Δ 4 and Δ 334 deletion vectors, concerning which only 32 of 46 clones analyzed were targeted on the basis of Southern blots (Table 1), and therefore 6TG resistance overestimated the targeting frequencies of these vectors, and spontaneous *HPRT* mutations may have accounted for some of the untargeted, 6TG-resistant colonies.

The structures of the *HPRT* loci were also assessed by Southern blots, and the 6TG-resistant clones infected with deletion vectors did not have the patterns expected for accurate targeting. As shown in Fig. 1C and D, representing fibroblast and HT-1080 clones, respectively, digestion with *Hind*III and *Ahd*I, and probing with right homology arm sequences, produced the expected 2.3-kb fragment in wild-type cells, a 3.6-kb fragment in cells targeted with the PN vector, and a 7.9-kb fragment in cells targeted with the PN(Δ 334) vector. However, cells targeted with the Δ 334 deletion vector had variable and multiple fragment sizes, whereas a single 6.6-kb fragment was expected. Although it was too complex to determine the actual structure of these *HPRT* loci, the pattern clearly showed that the targeting was inaccurate in these clones. These results (as well as those from blots with other restriction enzymes and probes not shown) are summarized in Table 1, where accurate targeting was observed in 12 of 12 clones targeted with the PN insertion vector, 22 of 23 clones targeted with the PN(Δ 334) insertion/deletion vector, but only 6 of 13 clones targeted (6 of 23 clones analyzed) with the Δ 334 deletion vector.

A similar phenomenon was observed for the +4 and Δ 4 vectors. Insertion of 4 bp was accurate in 40 of 41 6TG-resistant clones infected with AAV-HPe3(+4), but removal of 4 bp was accurate in only 6 of 23 6TG-resistant clones infected with AAV-HPe3(Δ 4) (Table 1). As observed for larger deletions, Southern blots showed that the 4-bp deletion vector targeted inaccurately (Fig. 2A). In these digests, the enzyme *Bso*BI was used, because the +4 and Δ 4 mutations both

FIG. 2. Structures of *HPRT* loci targeted with AAV-HPe3(+4) and AAV-HPe3(Δ4). (A) Southern blots of representative 6TG-resistant fibroblast clones transduced with the AAV-HPe3(+4) or AAV-HPe3(Δ4) targeting vectors and portrayed as in Fig. 1C. Genomic DNA was digested with *Bso*BI and probed as in Fig. 1. (B–G) Possible recombination events that would produce targeting at the *HPRT* locus are shown with the AAV terminal repeats as hatched boxes, homology arms as gray boxes, *HPRT* exons as black boxes; B, *Bso*BI sites. Recombination crossovers are shown as crossed lines. Predicted sizes of fragments on the Southern blot are indicated.

destroy a *Bso*BI site present in the wild-type locus, allowing one to detect targeting events. Southern blots showed the expected 2.0-kb fragment in wild-type cells, the expected 5.1-kb fragment in cells targeted with the +4 vector, but unexpected patterns in cells targeted with the Δ4 vector.

These unexpected Southern blot patterns fell into different classes, and could be explained by homologous recombination events that occurred with vector genome dimers, a phenomenon previously observed in a small percentage of clones targeted at the *COL1A1* or *COL1A2* locus (Chamberlain *et al.*, 2004, 2008). If linear, episomal vector genomes ligate to each other via their terminal repeats, they can then re-

combine with chromosomal target loci at their distal homology arms, effectively converting a monomer deletion vector into a dimer insertion vector. This is diagrammed in Fig. 2, where recombination with a vector monomer is shown in Fig. 2B, and recombination with head-to-tail vector dimers is shown in Fig. 2C–E. Depending on where the recombination crossovers occur within a vector genome dimer, a different restriction pattern is generated, with potential incorporation of a single left Δ4 mutation, a single right Δ4 mutation, or both left and right Δ4 mutations into the chromosome. This can be distinguished by *Bso*BI digestion, also taking into account the *Bso*BI sites present in the AAV ter-

minimal repeats that would be incorporated into the chromosome. On the basis of the Southern blot patterns observed in Fig. 2A, we can conclude that $\Delta 4$ -targeted clones 1, 2, 3, and 5 represent dimer recombination with the crossover pattern observed in Fig. 2C, $\Delta 4$ -targeted clone 4 represents dimer recombination with the crossover pattern shown in Fig. 2D, and $\Delta 4$ -targeted clone 6 represents accurate recombination with a monomer vector genome as occurred with +4-targeted clones. An alternative explanation for some of these patterns is recombination with a circularized vector genome (Fig. 2F and G), which allows a single crossover to create an insertion at the target locus. However, this is only consistent with $\Delta 4$ -targeted clone 4, which could be due to the recombination event shown in Fig. 2F. More complex scenarios are also possible with circularized vector multimers, but they all effectively convert a deletion vector into an insertion vector, and cannot necessarily be distinguished from recombination with linear vector multimers. Finally, it is possible that the chromosome could recombine with head-to-head or tail-to-tail vector dimers, but the inverted sequences present in the recombination intermediate may not resolve appropriately, and could instead lead to a loss of X chromosome sequences telomeric to the crossover point that would presumably be lethal.

Alkaline phosphatase gene targeting

To determine whether the preference for insertions during gene targeting can occur at other chromosomal target sites, we used an alkaline phosphatase (AP) reporter gene system. In these experiments, mutant AP genes are introduced into cells by retroviral vectors, the mutations are subsequently corrected by an AAV gene-targeting vector, and the gene-targeting frequencies are scored by staining infected cultures for AP expression (Hirata and Russell, 2000). A single vector stock is used to correct different mutations, so potential differences in the biological activity of stock preparations are not an issue. In addition, a polyclonal population of cells is used, containing more than 10^4 distinct vector proviruses located at different chromosomal positions, so gene-targeting frequencies are measured on a genome-wide basis.

We used this assay to compare the gene-targeting frequencies of AP genes with a 1-bp deletion or a 1-bp insertion located at bp 375 of the AP reading frame (Fig. 3A). Polyclonal populations of immortalized human fibroblasts containing each of these mutant AP genes were infected with the AAV-5'APBss targeting vector, which contains a non-functional, truncated AP gene with wild-type sequence at the bp 375 mutation site. The number of AP⁺ foci was 14.7-fold higher in the cells containing the 1-bp deletion in AP. No AP⁺ foci were observed in cells that did not receive the vector. This confirms that insertions are introduced at higher frequencies than deletions, even when the mutations are as small as a single base pair.

Gene targeting with transfected plasmid constructs

We tested whether the preference for insertions was a more general phenomenon by using the more conventional approach of transfected plasmid constructs. Because the targeting frequencies in human cells are typically 3-4 logs lower with transfection as opposed to transduction with AAV vec-

FIG. 3. Correction of AP mutations by gene targeting. **(A)** The structure of retroviral MLV-LAPVH-based vectors are shown with murine leukemia virus long terminal repeats (LTRs), AP gene, internal ribosome entry site (IRES), hygromycin resistance gene, and the +1, $\Delta 1$, and wild-type (WT) sequences surrounding the bp 375 mutation site indicated above the AP gene. The AAV-5'APBss targeting vector is shown with hatched boxes representing the inverted terminal repeats, homology to the MLV-LAPVH elements indicated, a 5' portion of the wild-type AP gene, and the BssHII sites used to prepare the targeting vector. **(B)** Gene-targeting frequencies of immortalized fibroblasts containing +1 (solid column) or $\Delta 1$ (open column) AP mutations and transduced with AAV-5'APBss are shown as the number of AP⁺ foci obtained per 10^5 infected cells (mean $\pm$ standard deviation, $n = 3$).

tors, we developed a different system for measuring targeting frequencies. HT-1080 cells were used because they are more easily transfected than human fibroblasts, and we decided to correct *HPRT* mutations rather than inactivate *HPRT*. This allowed us to select for *HPRT*⁺ cells with HAT medium soon after infection, and avoid the prolonged culture period required to eliminate *HPRT* protein before selecting for *HPRT*[−] cells with 6TG. Because we have to transfect more than 10^7 cells in order to obtain a single gene-targeting event, this culture period would result in an impractical cell expansion.

Four different HT-1080 clones were obtained by AAV-mediated gene targeting and confirmed by Southern blots as shown in Fig. 1: two each with the $\Delta 4$ and +4 mutations in exon 3 of *HPRT*. The selected clones had single copies of the *HPRT* target loci, including those with the $\Delta 4$ mutations (multimer-targeted clones were excluded). Each of these clones was infected with the AAV-HPe3 vector (which contains wild-type *HPRT* sequences surrounding exon 3), or transfected with an analogous, linearized plasmid fragment of wild-type *HPRT* sequences (Fig. 4A). Although the AAV vector targeted at much higher frequencies, as observed previously (Russell and Hirata, 1998), both AAV-mediated and plasmid-mediated gene targeting were more efficient when inserting as opposed to deleting 4 bp (25- to 28-fold and 11- to 23-fold higher frequencies, respectively; Fig. 4B). Thus the

preference for inserting nucleotides during gene targeting appears to be a general phenomenon in human cells, and not a feature unique to AAV.

Discussion

In these experiments we have established that insertions are introduced at higher frequencies than deletions during gene targeting. The phenomenon holds true for insertions ranging from 1 to 1332 bp, and deletions ranging from 1 to 334 bp (larger insertions and deletions were not tested), for different target genes and loci, and for both AAV-mediated and plasmid-mediated gene targeting. The magnitude of the effect can vary, but in several experiments there was a more than 10-fold preference for insertions. In some cases this preference was underestimated, because Southern blots showed that the deletion vectors often recombined with the targeted loci as multimers and actually produced insertion mutations. At this point the data are too preliminary to determine whether larger insertions are preferentially introduced relative to smaller insertions, but this may have been true for the 1332-bp insertion in AAV-HPe3PN as compared with the 4-bp insertion in AAVHPe3(+4) (Fig. 1).

Although gene targeting is a widely used experimental method, the targeting frequencies of insertion and deletion constructs are rarely compared, and we are not aware of a prior report describing preferential targeting by insertion vectors. In one particularly relevant paper, Zhang and coworkers compared the targeting frequencies of *hprt* deletion vectors in mouse embryonic stem cells and noted similar frequencies in vectors designed to insert 1.7 kb, or delete up to 19.2 kb (Zhang *et al.*, 1994). However, the "deletion" vectors tested also inserted a PGK-*neo* cassette, so they were analogous to the combined "insertion/deletion" vector we used in our experiments, which targeted at intermediate frequencies. It should be noted that Southern blot analysis of clones in that study, as well as those targeted by "replacement" vectors in other experiments (Hasty *et al.*, 1991), have observed multimerization of targeting constructs, similar to what we report here. On the basis of our findings, one might also expect that studies of spontaneous mutagenesis would have noted preferential incorporation of insertion mutations rather than deletion mutations. However, prior studies are inconclusive in this regard, with some reporting a repair bias toward insertions (Taghian *et al.*, 1998; Twerdi *et al.*, 1999) and others toward deletions (Weiss and Wilson, 1987; Bill *et al.*, 2001). Additional experiments will be required to determine whether this is due to differences in the species, cell lines, or assays used to measure gene targeting and DNA repair.

Our findings suggest that cells process different types of recombination intermediates as heteroduplexes that form between the vector genome and the chromosome. This is presumably the only time the cell can distinguish between insertion and deletion mutations, especially the subtle difference between a 1-bp deletion and insertion. These mutations will form unpaired loops in the heteroduplex that can be recognized by mismatch repair or related proteins (Jiricny, 2006). This could lead to preferential unwinding of heteroduplexes before recombination when the deletion is in the AAV strand, and/or preferential repair of deletions after recombination occurs. A similar heteroduplex-dependent pro-

FIG. 4. Correction of *HPRT* mutations with transfected plasmid constructs or AAV vectors. **(A)** Structures of the linear plasmid fragment used for targeting, the AAV-HPe3 targeting vector, and the *HPRT* target loci containing +4 or Δ4 mutations in exon 3 are shown with the AAV terminal repeats as hatched boxes, homology arms as gray boxes, and *HPRT* exons as black boxes. **(B)** The fraction of HAT-resistant colonies obtained from untransfected, untransduced cells, cells transduced with AAV-HPe3, and cells transfected with the pHPe2/3 plasmid fragment is shown, with each column representing one experimental value. Two clones (c1 and c2), containing each type of mutation as indicated, were tested. Asterisks indicate that no HAT-resistant colonies were obtained and the columns below the asterisks show the maximal possible value given the sensitivity of the experiment. The lower cell number used when infecting with AAV reduced the sensitivity of the experiment with the +4-c2 clone, and thus the true targeting frequency could be lower in this case where a deletion was being introduced.

cess must also occur during recombination between double-stranded plasmid constructs transfected into cells and chromosomal target loci. This could represent a normal process of DNA repair/recombination that preserves insertions recognized during heteroduplex formation between chromosomes. This would prevent the loss of genetic information over time, but could also lead to the deleterious incorporation of additional nucleotides, for example, during the expansion of disease-causing trinucleotide repeats (Pearson *et al.*, 2005).

Our results also have implications for the design of gene-targeting vectors. If possible, one should use vectors that create simple insertions without removing any bases. If a deletion is required, one should use an insertion/deletion vector and then, if necessary, remove the insertion later on by Cre-mediated recombination (Sauer and Henderson, 1989) or

some other approach. Even if low targeting frequencies are not a problem due to powerful selection methods, deletion vectors will still tend to target inaccurately and should therefore be avoided. It may also be possible to improve overall targeting rates by introducing silent insertions along with whatever mutation is desired. This could be especially important during therapeutic gene targeting, when there is little room for flexibility in experimental design, and targeting rates are critical.

Acknowledgments

The authors thank Cong Xu and Richard Newton for technical assistance, and Marie Anderson for the pLAPIH plasmid. This work was supported by grants from the U.S. National Institutes of Health.

Author Disclosure Statement

No competing financial interests exist.

References

- Bill, C.A., Taghian, D.G., Duran, W.A., and Nickoloff, J.A. (2001). Repair bias of large loop mismatches during recombination in mammalian cells depends on loop length and structure. *Mutat. Res.* 485, 255–265.
- Chamberlain, J.R., Schwarze, U., Wang, P., Hirata, R.K., Hankenson, K.D., Pace, J.M., Underwood, R.A., Song, K.M., Sussman, M., Byers, P.H., and Russell, D.W. (2004). Gene targeting in stem cells from individuals with osteogenesis imperfecta. *Science* 303, 1198–1201.
- Chamberlain, J.R., Deyle, D.R., Schwarze, U., Wang, P., Hirata, R.K., Li, Y., Byers, P.H., and Russell, D.W. (2008). Gene targeting of mutant *COL1A2* alleles in mesenchymal stem cells from individuals with osteogenesis imperfecta. *Mol. Ther.* 16, 187–193.
- Fields Berry, S.C., Halliday, A.L., and Cepko, C.L. (1992). A recombinant retrovirus encoding alkaline phosphatase confirms clonal boundary assignment in lineage analysis of murine retina. *Proc. Natl. Acad. Sci. U.S.A.* 89, 693–697.
- Hasty, P., Rivera-Pérez, J., Chang, C., and Bradley, A. (1991). Target frequency and integration pattern for insertion and replacement vectors in embryonic stem cells. *Mol. Cell. Biol.* 11, 4509–4517.
- Hendrie, P.C., and Russell, D.W. (2005). Gene targeting with viral vectors. *Mol. Ther.* 12, 9–17.
- Hirata, R.K., and Russell, D.W. (2000). Design and packaging of adeno-associated virus gene targeting vectors. *J. Virol.* 74, 4612–4620.
- Hirata, R., Chamberlain, J., Dong, R., and Russell, D.W. (2002). Targeted transgene insertion into human chromosomes by adeno-associated virus vectors. *Nat. Biotechnol.* 20, 735–738.
- Inoue, N., Hirata, R.K., and Russell, D.W. (1999). High-fidelity correction of mutations at multiple chromosomal positions by adeno-associated virus vectors. *J. Virol.* 73, 7376–7380.
- Inoue, N., Dong, R., Hirata, R.K., and Russell, D.W. (2001). Introduction of single base substitutions at homologous chromosomal sequences by adeno-associated virus vectors. *Mol. Ther.* 3, 526–530.
- Jiricny, J. (2006). The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell. Biol.* 7, 335–346.
- Miller, A.D., Garcia, J.V., von Suhr, N., Lynch, C.M., Wilson, C., and Eiden, M.V. (1991). Construction and properties of retrovirus packaging cells based on gibbon ape leukemia virus. *J. Virol.* 65, 2220–2224.
- Miller, D.G., Edwards, R.H., and Miller, A.D. (1994). Cloning of the cellular receptor for amphotropic murine retroviruses reveals homology to that for gibbon ape leukemia virus. *Proc. Natl. Acad. Sci. U.S.A.* 91, 78–82.
- Patel, P.I., Framson, P.E., Caskey, C.T., and Chinault, A.C. (1986). Fine structure of the human hypoxanthine phosphoribosyltransferase gene. *Mol. Cell. Biol.* 6, 393–403.
- Pearson, C.E., Nichol Edamura, K., and Cleary, J.D. (2005). Repeat instability: Mechanisms of dynamic mutations. *Nat. Rev. Genet.* 6, 729–742.
- Rasheed, S., Nelson Rees, W.A., Toth, E.M., Arnstein, P., and Gardner, M.B. (1974). Characterization of a newly derived human sarcoma cell line (HT-1080). *Cancer* 33, 1027–1033.
- Russell, D.W., and Hirata, R.K. (1998). Human gene targeting by viral vectors. *Nat. Genet.* 18, 325–330.
- Sauer, B., and Henderson, N. (1989). Cre-stimulated recombination at *loxP*-containing DNA sequences placed into the mammalian genome. *Nucleic Acids Res.* 17, 147–161.
- Taghian, D.G., Hough, H., and Nickoloff, J.A. (1998). Biased short tract repair of palindromic loop mismatches in mammalian cells. *Genetics* 148, 1257–1268.
- Thomas, K.R., and Capecchi, M.R. (1986). Introduction of homologous DNA sequences into mammalian cells induces mutations in the cognate gene. *Nature* 324, 34–38.
- Twerdi, C.D., Boyer, J.C., and Farber, R.A. (1999). Relative rates of insertion and deletion mutations in a microsatellite sequence in cultured cells. *Proc. Natl. Acad. Sci. U.S.A.* 96, 2875–2879.
- Valancius, V., and Smithies, O. (1991). Double-strand gap repair in a mammalian gene targeting reaction. *Mol. Cell. Biol.* 11, 4389–4397.
- Vasileva, A., and Jessberger, R. (2005). Precise hit: Adeno-associated virus in gene targeting. *Nat. Rev. Microbiol.* 3, 837–847.
- Weiss, U., and Wilson, J.H. (1987). Repair of single-stranded loops in heteroduplex DNA transfected into mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* 84, 1619–1623.
- Zhang, H., Hasty, P., and Bradley, A. (1994). Targeting frequency for deletion vectors in embryonic stem cells. *Mol. Cell. Biol.* 14, 2404–2410.

Address reprint requests to:
Dr. David W. Russell
Mailstop 357720
University of Washington
1705 NE Pacific Street
Seattle, WA 98195-7720

E-mail: drussell@u.washington

Received for publication June 9, 2008; accepted after revision July 30, 2008.

Published online: August 28, 2008.